

Db	906	NNSEILLYKCLNQSREK	922
DR	EMBL: AJ294725; CAC24573..1; - .		
DR	HSSP; Q9KWT7; 1HOM.		
DR	InterPro; IPR001572; RNA_B01_B.		
DR	PFam; PF04163; RNA_pol_B01.		
DR	PFam; PF04161; RNA_pol_Rpb2_1..1.		
DR	PFam; PF04165; RNA_pol_Rpb2_2..1.		
DR	PFam; PF04162; RNA_pol_Rpb2_6..2.		
DR	PFam; PF00062; RNA_pol_Rpb2_7..1.		
DR	PROSITE; PS01166; RNA_POL_BETA_1..1.		
DR	Transerase; DNA-directed RNA polymerase; Transcript; Chloroplast.		
GW	SEQUENCE 1076 AA; direct; 5D23D48E8B3EB2D76 CRC64;		
Q9	SEQUENCE 124362 MW; 5D23D48E8B3EB2D76 CRC64;		
Dy	Query Match Score 219; DB 1; Length 1076;		
Dy	Best Local Similarity 21.0%; Pred. No. 0.014;		
Dy	Matches 180; Conservative 145; Mismatches 276; Indels 256; Gaps 49;		
Dy	20 SLKKDTGCVHHHQHNEENESIKPKSSPMI-----DRNISTIRD----FENKDLKK 64		
Dy	180 SCKKKTLSLNKNNSYKNSFKISKKKTBIPTINTLTSKENNTKTINNLYRFLNSDYN 239		
Dy	65 L-----IKKFREVDFTSETGKRMEE---YDYKYDDKKNIIAYDDGTDLVEYTEKL 113		
Dy	240 LGDTGRFKINKKIKYKTEFTNNKIKLMPDFLGJENYMIK-----NINIKSNKI 289		
Dy	114 DEIKSKTYGVLS-SPS-----KGDFHEFLIGKISNV--SKNAKVYYGNNYKSIEIKA 160		
Dy	290 DQJRNKI--VLSVGELMNQNKFNNIKQDLYTKI.EKINKFEOQKKOBERYKNERKEEKIKI 347		
Dy	161 TK-YDPHSKTMFDL--YANINDIVDGLAFAQDMR-LFVKONDQKAEKIKRMPPEKIK-E 215		
Dy	348 NCTYFTNSKNLTDNICKPITNPLSOLL--NDLNPLSELTHKRKTSSTLGIGIEKKNKA 404		
Dy	216 TKSBEYPVSSYGNV--TELGBEDSLSKNRPDNLTIKMEGKIVSDSERQYLLKDNTILRKG 273		
Dy	405 TKIREINSHYGRICPLETSEG--KN-----AGVLVS-----LAKDIRINKHG 445		
Dy	274 YALKVTYTNPGKTDMLBEGNYY--SKB-----DIAKIQKANENLRALESETTIYAD 321		
Dy	446 F-IESPFKVKIGENIKGNKGFISSNEENKN-KIAPDILKNEFLNKNYGVKNKNEYD 504		
Dy	322 SRNVEDGRSTQSVLMSALDGPNVIRYQVTF-KMDNGKEAI-----DKDGNL 367		
Dy	505 S-----YKSTINFISTSTDQFNSIGTGJLIPFLEHNDDANRVIMGATMQKOALLLNKEPSL 558		
Dy	368 VTDSSKUFLVFGDKEYTGEDIKFENVEAIKDSMLFDTPKP-----NLSMDK 415		
Dy	559 I-ETGRELLINDSK-----STI LAKESGTVIYSSSKKIIIOQSKSKFKQNINLFK 609		
Dy	416 NYENP--SKSNKLYVRAAPEFYLRGKLSDKGGPNWELRVNESTVDNYLIYGDIHIDNTRDF 473		
Dy	610 KPFKRLKRNKNCIKKHKHTIYL-----LNUDEK--NNNTY-----ORTS 648		
Dy	474 NTKLN--VKDGDIMDWGMKDYKANGFPDKVTMDGNYL--QTGYSDNA---KAVG 523		
Dy	649 IVKNDWIKKGQIAEG-----IGSLNGNLCLGKNIILVYLSWEGCFNFEDAI 696		
Dy	524 VHYQFLDVNPVKPVNIDPKGNTSIEYADGKSVTF-----NINDKRANGFDGEBIQEQ 574		
Dy	697 ISERLRNEDILTSIHVKCKSKFLINHKEKKEETKNIPIDKUKVNKLNNNG---IIKI 752		
Dy	575 HYINGKEYTSFNDIKQIIDIKTLNIKI-----VVKDGPAR-NTTV---KEFLINKDTGEVS 625		
Dy	753 GSKingkeil---GIIKKRILINYEVELIYEILSEYERKN-SVLSPRNLV-----GIVt 803		
Dy	626 ELKPHRVTVTIG-----NGKEMSSTIVSEEDFLPVYKGELKGYQFD 668		
Dy	804 NTIKHLKNCYQIEIYIETKKQIGDKLSGHGNKGSVISK---IIP-----VDMYLPD 856		
Dy	669 GNBIS-----GPFEGKDAGY-----INLSKOTFIEPVFKLBIERKGKQFDFVSKK 718		
Dy	857 GTFIDILINPLGJPSRANVQGYIECLNL-----AIIKLKERYKIQP-FDEYQK 905		
Dy	719 DNPVO-NHSOLNESHRK 734		
Qy	RESULT 2 MSP1-PLAFK STANDARD; PRT; 1630 AA. ID MSP1-PLAFK SEQUENCE FROM N.A. SEQUENCE: MSPL-PLAFK MEDLINE=86136024; PubMed=3004972; RX Mackay M., Goman M., Bone N., Hyde J.E., Scaife J., Certa U., RA Stunnenberg H., Bujard H.; RT "Polymorphism of the precursor for the major surface antigens of Plasmodium falciparum merozoites: studies at the genetic level.", RT Plasmodium falciparum (isolate Kl / Thailand); Apicomplexa; Haemosporida; Plasmoidium. DB (PMMSA) (P190). GN MSP-1. OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmoidium. NCBI_TaxID=5839; RN [1] RN SEQUENCE FROM N.A. RN Submitted (JUN-1995) to the EMBL/GenBank/DB/DBJ databases. RN REVISIONS, SEQUENCE FROM N.A. RN Pan W., Toille R., Bujard H.; RN Submitted (JUN-1995) to the EMBL/GenBank/DB/DBJ databases. RN SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor RN (Potential). RN PTM: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 kDa. RN KDA AND 19 kDa ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT. RN [2] RN CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation of the European Bioinformatics Institute. There are no restrictions on use by non-profit institutions as long as its content is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb-sib.ch/and/or send an email to license@isb-sib.ch). RN CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation of the European Bioinformatics Institute. There are no restrictions on use by non-profit institutions as long as its content is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb-sib.ch/and/or send an email to license@isb-sib.ch). RN CC DR EMBL: X03371; CAA27070_1; -. RN DR InterPro; IPR006209; EGF-like. RN DR Pfam; PF00008; EGFL_1. RN KW Malaria; Merozoite; Polypeptide; Repeat; Signal; Glycoprotein; Transmembrane; GPI-anchor. RN KW SIGNAL 1 19 POTENTIAL. MEROZOITE SURFACE PROTEIN 1. RN FT CHAIN 20 1630 TRIPETIDE SG (TP) REPEAT. RN FT DOMAIN 67 84 MEMBRANE ANCHOR. RN FT TRANSMEM 1614 1630 RN FT CARBOHYD 97 97 N-LINKED (GLCNAC. . .) RN FT CARBOHYD 259 259 N-LINKED (GLCNAC. . .) RN FT CARBOHYD 755 755 N-LINKED (GLCNAC. . .) RN FT CARBOHYD 759 759 N-LINKED (GLCNAC. . .) RN FT CARBOHYD 774 774 N-LINKED (GLCNAC. . .) RN FT CARBOHYD 835 835 N-LINKED (GLCNAC. . .) RN FT CARBOHYD 911 911 N-LINKED (GLCNAC. . .) RN FT CARBOHYD 955 955 N-LINKED (GLCNAC. . .) RN FT CARBOHYD 1049 1049 N-LINKED (GLCNAC. . .) RN FT CARBOHYD 1156 1156 N-LINKED (GLCNAC. . .) RN FT CARBOHYD 1165 1165 N-LINKED (GLCNAC. . .) RN FT CARBOHYD 1436 1436 N-LINKED (GLCNAC. . .) RN FT CARBOHYD 1517 1517 N-LINKED (GLCNAC. . .) RN SQ SEQUENCE 1630 AA; 187289 MW; AUBDBC3CE04A6322 CRC64;	5.4%; Score 218; DB 1; Best Local Similarity 18.6%; Pred. No. 0.025; Matches 167; Conservative 146; Mismatches 340; Indels 228; Gap 23 KDTTG-VERHHQENENSIK-----EKSFTIDRNISTIRFENKQLKLKKKFREV	

Db	223	KDNVGRMEDYIKKNNKTINENILLEBESKTKIDKRNATKEE-----KKLYQAQY 274
Qy	76	FTSETGKRMEEYDYKYYDDKGKNNIAYDDGTDLLEYETKLDEIKSKKIYGVLSPLSKDGHFPEIL 135
Qy	275	DLSYNGKOLEE-----AHLNLSV-----LE-----KRIDLK-----:-----KNEINKELL 310
Db	136	GKISNVSKRAVKVYGGNYKSIEIKATKDFH-----SKMTFDIYANINDIVGLAFA 188
Qy	311	DCKNBTKNPPPANSNTPTNDLDRKKKIEHEKEIKAETIKENIDSFLFDPL-----364
Db	189	GDMRLFLVKONDOKKAETIKRMPKEIKTKSEYP-----YVSSYGNV-----IELGE-GDLSK 239
Qy	365	-ELLYYLREKN-KNIDISAKVETKESTEPNPYNGVTPLSYNDINNALNEINSFGPLI- 421
Db	240	NKPBDNLTKMBSGKTYSDSEKQYL-LKDNTIL-----RKGYALKVUTTYNPGKTDML 289
Qy	422	-NPDPDTKEPSKNTYTNDNERKFKFINIKEKEKKEKIESDRSKSLSN-----DIT 476
Db	290	EG-----NGVYSKE-----DIAKTOKA-----NPNURALSETTIYASDRNVEYDGRSTQSV 334
Qy	477	KEYKLNLNEYDSKRFNNNIDLTNFKOMMGKRYSYKVEKLTHHTNASYSAYENKHNLEKLTK 536
Qy	335	LMSALDFG---NII-----RYQVTFKNDXGEALDKDGNLTVTDSKVLFGKODDKEYT 385
Db	537	ALKYMDYSLRNIVBEKLYKKNLISKINELTIVE-----NIKODBEOLF-----EKIT 589
Qy	386	GEDKFVNVEAIKEDGSMFLIDTCKPVNLSMDKMYFNPKSNKSIYVNRNPEFYLRGKISDKGFF 445
Db	590	KDENKRPDEKLEKIVQVQKVLL--MNKIDELLKQTJLKNYE--LGKHNHYENSY 644
Qy	446	NWELRVNESVVDNYLLYGDHLIDNPRDFNKL-----NVK-----480
Db	645	KQENKQE---PYYLIVLKEIDKUKVEMPKVESLINEEKKNIKTEQQSDNSEPSTEGEI 700
Qy	481	------DGDIMDWGMKDDYK-----ANGFPDKV 501
Db	701	TGQATTPKPCQGASMLEGDSVAQQBQKQDAPPVPVPAKAVQVTPPPAVPNKBNV 760
Qy	502	TDMGGVVYQLTGSDINA----KAVGHHYQFLDYNVKPEYNIDPKGNTSLEYADGKSV 556
Db	761	SKLD---YLEKLYEFPLNTSYTICHKYTLVSHSTMNERKILQKQYKITKEESEKLUSSCDPDLI 817
Qy	557	FNIN-----DKRANNFDG---EIQBHQHYINGKEYTSFNDIKOIIDKTLNIKIVV 603
Db	818	FNTQNNIPYMSPMFSLNSLQSLPEMEIYKEMVCNLYKLKDNDKIKNLLEAKVYSTSV 877
Qy	604	KDPARTNTVKEFLNKDGTGEYSELKPHRVTVTIONCKEMSSTIVSEDFLPVYKGPLEK 663
Db	878	KTLUSSSSMOPSLTIEQDKPEVSANDTSHSTNLNSLKLFNITLSS-----LGRKNN 928
Qy	664	GYQFDGWLSGFEGK-----KDAFYVNLSKOTPIKPVFKKIEKKEEEENKPTF-D 713
Db	929	IYQ---ELJGQKSENFYEKILKUSDFTYNESETNFVKSCKADDLNSLNDSESKRKLLEED 984
Qy	714	VSKKDDNPOVNHSQLNESHRKEDLOREEHSGKSDSTKDVTATVLKNNISSK-STTNNP 771
Db	985	INKLKTQLQSFDLNKYKUKLKERLFDKKKTUVGRKMQIKLTLKEQLESKLNSLNNP 1043
RESULT 3		
	MSPL_PLAFW	STANDARD;
	ID MSPL_PLAFW	PRT: 1639 AA.
	AC P04933;	(Rel. 05, Created)
	DT 13-AUG-1987	(Rel. 33, Last sequence update)
	DT 01-FEB-1996	(Rel. 42, Last annotation update)
	DT 15-SEP-2003	(Rel. 42, Last annotation update)
	DB Merozoite surface protein 1 precursor (Merozoite surface antigens)	
	DE (PMMSA) (P195).	
	GN MSP-1.	
	OS Plasmodium falciparum (isolate Wellcome).	
	OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.	
	ON NCBI_TAXID=5848;	
	[1]	

SEQUENCE FROM N.A.			
RX MEDLINE=86014355; PubMed=2995820;			
RA Holder A.A.; Lockyer M.J.; Odink K.G.; Sandhu J.S.; Riveros-Moreno V.,			
RA Nicholas S.C.; Hillman Y.; Davy L.S.; Tizard M.L.V.; Schwarz R.T.,			
RA Freeman R.R.;			
RT PRIMARY STRUCTURE OF THE PRECURSOR TO THE THREE MAJOR SURFACE			
RT ANTIGENS OF PLASMODIUM FALCIPARUM MEROZOITES.;"			
RT Nature 317:270-273 (1985).			
RL [2]			
RN RN			
RP RP			
REVISIONS.			
RA Holder A.A.;			
RA Submitted (MAR-1991) to the EMBL/GenBank/DDBJ databases.			
CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor			
CC (potential).			
CC -1- PTM: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 kDa, 42			
CC kDa AND 19 kDa ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF			
CC MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.			
CC			
CC This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC between the Swiss Institute of Bioinformatics and the EMBL outstation			
CC at the European Bioinformatics Institute. There are no restrictions on			
CC use by non-profit institutions as long as its content is in no way			
CC modified and this statement is not removed. Usage by and for commercial			
CC entities requires a licence agreement (see http://www.ebi.ac.uk/announce or send an email to license@ebi.ac.uk).			
CC			
CC EMBL; X02919; CAA26676; 1; -.			
DR DR			
PIR; A24594; A24594.			
PTR; S05603; S05603.			
PDB; 1CEU; 2B-MY-99.			
DR DR			
InterPro; IPR006209; BGF-like.			
DR DR			
Pfam; PF00008; EGF_1.			
KW Malaria; Merozoite; Polyprotein; Repeat; Signal; Glycoprotein;			
KW Transmembrane; GPI-anchor; 3D-structure.			
CC			
FT SIGNAL 1 19 POTENTIAL.			
FT CHAIN 20 1639 MEROZOITE SURFACE PROTEIN 1.			
FT CARBOHYD 116 116 N-LINKED (GLCNAC. . .) (POTENTIAL).			
FT CARBOHYD 268 268 N-LINKED (GLCNAC. . .) (POTENTIAL).			
FT CARBOHYD 764 764 N-LINKED (GLCNAC. . .) (POTENTIAL).			
FT CARBOHYD 768 768 N-LINKED (GLCNAC. . .) (POTENTIAL).			
FT CARBOHYD 783 783 N-LINKED (GLCNAC. . .) (POTENTIAL).			
FT CARBOHYD 844 844 N-LINKED (GLCNAC. . .) (POTENTIAL).			
FT CARBOHYD 920 920 N-LINKED (GLCNAC. . .) (POTENTIAL).			
FT CARBOHYD 964 964 N-LINKED (GLCNAC. . .) (POTENTIAL).			
FT CARBOHYD 1058 1058 N-LINKED (GLCNAC. . .) (POTENTIAL).			
FT CARBOHYD 1165 1165 N-LINKED (GLCNAC. . .) (POTENTIAL).			
FT CARBOHYD 1174 1174 N-LINKED (GLCNAC. . .) (POTENTIAL).			
FT CARBOHYD 1445 1445 N-LINKED (GLCNAC. . .) (POTENTIAL).			
FT CARBOHYD 1526 1526 N-LINKED (GLCNAC. . .) (POTENTIAL).			
SEQUENCE 1639 AA; 187618 MW; 2C255B8616C87F6E CRC64;			
SQ			
Query Match 5.4%; Score 218; DB 1; Length 1639;			
Best Local Similarity 18.6%; Pred. No. 0.025;			
Matches 167; Conservative 164; Mismatches 340; Indels 228; Gaps 3			
Qy 23 KDTG-VEHHHOENESSIK----EKSSEFTIDRNISTIRDENKLKKLKKKFREVDD 75			
Db 232 KDNVGRMEDYKDDKGNTIAYDDGTOLLEYETRKLDKSKYQVLS9SKDGHFBIL 13			
Qy 76 FTSETOKRMEEEDYKDDKGNTIAYDDGTOLLEYETRKLDKSKYQVLS9SKDGHFBIL 13			
Db 284 DLSLYNKQLEE----AHNLISV----LE----KRDTLK-----KNEKELL 31			
Qy 136 GRISNVSKNAKYGVGNVYKSTBIKAKYDFH-----SKMTFDLYANINDIVDGLAFA 18			
Db 320 DKINEKNPPPANSQNTPTNLLDKQKKEKEKEIKEIARTIKENIDSLFTDP-----37			
Qy 189 GDMRLEFKNDQDKKARIKIRMPKEKIKETKSFYP---YVSSYGNV---IEGE-GDLSK 23			
Db 374 -ELEYTLREKKN-KNIDISAKVETKESTEPMNGVTPYPLSYNDINNALNEINSFGSDLI- 43			
Qy 240 NKPDLNITKMEGKQIYSDSEKQQL-IKDNL-----RGKYLAKUTTYNPKGTDML 28			

KW	Sporulation; Coiled coil.	COILED COIL (POTENTIAL).
01-OCT-1996 (Rel. 3.4, Created)	FT DOMAIN 199 785	COILED COIL (POTENTIAL).
01-OCT-1996 (Rel. 3.4, Last sequence update)	FT DOMAIN 804 1235	COILED COIL (POTENTIAL).
15-SEP-2003 (Rel. 4.2, Last annotation update)	FT DOMAIN 1320 1471	COILED COIL (POTENTIAL).
sp015 OR SPAC1F3_06.C.	FT DOMAIN 1481 1723	COILED COIL (POTENTIAL).
SQ SEQUENCE 1957 AA:	222785 MW:	3F480CA0617D9DA CRC64;
	Query Match 4.8%; Score 194; DB 1; Length 1957;	
[1] SEQUENCE FROM N.A., FUNCTION, SUBUNIT, AND SUBCELLULAR LOCATION.	Best Local Similarity 20.5%; Pred. No. 0.33; Mismatches 290; Index 238; Gaps 44; Matches 174; Conservative 147; MisMatches 290;	
	SEQUENCE-1972; MEDLINE=21848401; PubMed=11859360; MEDLINE=20107136; PubMed=10639340; SGouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S., Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A., Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G., Holroyd S., Hornsby T., Howarth S., Hunt S., Jagels K., James K., Jones L., Jones M., Leather S., McDonald S., McLean J., Mooney P., Moule S., Mongall K., Murphy L., Niblett D., Odell C., Oliver K., O'Neil S., Pearson D., Quail M.A., Rabinowitzch E., Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S., Skelton J., Simmonds M., Squares R., Stevens S., Walsh S.V., Warren T., Whitehead S., Taylor K., Taylor R.G., Tilvey A., Woodward J., Volkert G., Aert R., Robben J., Grimonprez B., Weltjens I., Vansbreels E., Rieger M., Schaefer M., Mueller-Auer S., Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H., Borzym K., Langner I., Beck A., Lehrach H., Reinhardt R., Pohl T.M., Beger P., Zimmermann W., Wedler H., Wambutt R., Purcell B., Goffeau A., Cadieu E., Dreano S., Gloux S., Lejaire V., Mottier S., Lucas M., Rochet M., Gaillard C., Tallada V.A., Garzon A., Thode G., Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J., Dominguez A., Revuelta J.M., Moreno S., Armstrong J.J., Forssburg S.L., Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J., Shpakovski G.V., Ussey D., Barrell B.G., Nurse P.; The genome sequence of <i>Schizosaccharomyces pombe</i> .; Nature 415:871-880(2002).	STRAIN-972; MEDLINE=21848401; PubMed=11859360; MEDLINE=20107136; PubMed=10639340; SGouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S., Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A., Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G., Holroyd S., Hornsby T., Howarth S., Hunt S., Jagels K., James K., Jones L., Jones M., Leather S., McDonald S., McLean J., Mooney P., Moule S., Mongall K., Murphy L., Niblett D., Odell C., Oliver K., O'Neil S., Pearson D., Quail M.A., Rabinowitzch E., Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S., Skelton J., Simmonds M., Squares R., Stevens S., Walsh S.V., Warren T., Whitehead S., Taylor K., Taylor R.G., Tilvey A., Woodward J., Volkert G., Aert R., Robben J., Grimonprez B., Weltjens I., Vansbreels E., Rieger M., Schaefer M., Mueller-Auer S., Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H., Borzym K., Langner I., Beck A., Lehrach H., Reinhardt R., Pohl T.M., Beger P., Zimmermann W., Wedler H., Wambutt R., Purcell B., Goffeau A., Cadieu E., Dreano S., Gloux S., Lejaire V., Mottier S., Lucas M., Rochet M., Gaillard C., Tallada V.A., Garzon A., Thode G., Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J., Dominguez A., Revuelta J.M., Moreno S., Armstrong J.J., Forssburg S.L., Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J., Shpakovski G.V., Ussey D., Barrell B.G., Nurse P.; The genome sequence of <i>Schizosaccharomyces pombe</i> .; Nature 415:871-880(2002).
	SEQUENCE OF 705-871 FROM N.A.	
	SEQUENCE-978 h90; MEDLINE=20233868; PubMed=10759889; Ding D.-Q., Tomita Y., Yamamoto A., Chikashige Y., Haraguchi T., Hiraoaka Y.; Large-scale screening of intracellular protein localization in living fission yeast cells by the use of a GFP-fusion genomic DNA library.; Genes Cells 5:169-190(2000).; -!- FUNCTION: Has a role in the initiation of spore membrane formation. -!- SUBUNIT: Monomer. -!- SUBCELLULAR LOCATION: Spindle pole body.	SEQUENCE-978 h90; MEDLINE=20233868; PubMed=10759889; Ding D.-Q., Tomita Y., Yamamoto A., Chikashige Y., Haraguchi T., Hiraoaka Y.; Large-scale screening of intracellular protein localization in living fission yeast cells by the use of a GFP-fusion genomic DNA library.; Genes Cells 5:169-190(2000).; -!- FUNCTION: Has a role in the initiation of spore membrane formation. -!- SUBUNIT: Monomer. -!- SUBCELLULAR LOCATION: Spindle pole body.
	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
	EMBL; AB027811; BA87115.1; -; PIR; T38077; T38077; GeneDB_Spombe; SPAC1F3_06C; -;	EMBL; AB027811; BA87115.1; -; PIR; T38077; T38077; GeneDB_Spombe; SPAC1F3_06C; -;
	Query 754 ATVDKNNI 762	Query 659 GELEYQYQFDGWEISGFEGKGDAGYVNLISKDTFKPVFKKIEE----KS 554
	Db 1629 NQVKDLNSI 1637	Db 1530 -----NNEYSLKLN-----DKDSITRDLSENIEQLNLLAEEKSAVRLS 1568
	Query 703 -KKEEE---NKPFDFVS -KKDNQPVNHSQNEHRSQKSQSDSTK-DVT 753	Query 703 -KKEEE---NKPFDFVS -KKDNQPVNHSQNEHRSQKSQSDSTK-DVT 753
	Db 1569 TEKESELOPNSRLADLEYTHKSQYESLGRSKLKAJSTBELQLAENERSLTRMLDLQ 1628	Db 1569 TEKESELOPNSRLADLEYTHKSQYESLGRSKLKAJSTBELQLAENERSLTRMLDLQ 1628
	Query 754 ATVDKNNI 762	Query 754 ATVDKNNI 762

RESULT 8

FOUND IN EUKARYOTIC NUCLEI: POLYMERASE I FOR THE RIBOSOMAL RNA PRECURSOR, POLYMERASE II FOR THE mRNA PRECURSOR, AND POLYMERASE III FOR 5S AND tRNA GENES
!- SIMILARITY: BELONGS TO THE RNA POLYMERASE BETA' CHAIN FAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation of the European Bioinformatics Institute. There are no restrictions on use by non-profit institutions as long as its content is not modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/ann.html>) or send an email to license@isb-sib.ch.

		RESULT 10	USO1 YEAST ^a	STANDARD;	PRT;	1790 AA.
RN		SEQUENCE FROM N.A.				
RP	ID-USO1	STRAIN=X-180-IA;				
RC	AC-P25386;	MEDLINE=91185402; PubMed=2010462;	(Rel. 22, Created)			
DT	01-MAY-1992	DT 01-MAY-1992 (Rel. 22, Last sequence update)				
DT	16-OCT-2001	DT 16-OCT-2001 (Rel. 40, Last annotation update)				
RX		Intracellular protein transport protein USO1.				
RA	USO1 OR INT1 OR YD1058W.					
RA	OS-GN	Saccharomyces cerevisiae (Baker's Yeast).				
RA	OC	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomyces;				
RA	OC	Saccharomyctaceae; Saccharomyces cerevisiae; Saccharomyces.				
OX	[1]	NCBI_TaxID=4932;				
RN		SEQUENCE OF 782-1790 FROM N.A.				
RP		SEQUENCE OF 782-1790 FROM N.A.				
RC		HOSTETTER M.K.; HERMAN D.J.; BENDEL C.M.; McCLELLAN M.; TAO N.,				
RA		"A cytoskeleton-related gene, uso1, is required for intracellular protein transport in Saccharomyces cerevisiae."				
RA		J. Cell Biol. 113:245-260(1991).				
RN		SEQUENCE OF 782-1790 FROM N.A.				
RP		HOSTETTER M.K.; HERMAN D.J.; BENDEL C.M.; McCLELLAN M.; TAO N.,				
RC		"A cytoskeleton-related gene, uso1, is required for intracellular protein transport in Saccharomyces cerevisiae."				
RA		J. Cell Biol. 113:245-260(1991).				

-1- CAUTION: REF 2 MISQUOTES THE GENE NAME AS "MPLI".

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to licensee@isb-sib.ch).

CC DR EMBL; I01992; AA034783; 1; -.
DR EMBL; X73541; CA051948; 1; -.
DR EMBL; 228320; CA082174; 1; -.
PDR; S38173; S38173.
DR SGD; S0001803; MUP1.
DR GO; GO:0005635; C:nuclear membrane; IDA.
DR GO; GO:0006006; P:protein-nucleus import; IDA.
KW Coiled coil; DNA repair.

FT DOMAIN 69 487 COILED COIL (POTENTIAL).
FT DOMAIN 531 1678 COILED COIL (POTENTIAL).
FT DOMAIN 1834 1866 COILED COIL (POTENTIAL).
FT CONFLICT 301 301 R -> A (IN REF. 1).
SQ SEQUENCE 1875 AA; 218455 MW;

Query Match Score 4.63; Best Local Similarity 19.4%; Pred. No. 0, 85 %; Length 1875;
Matches 173; Conservative 168; Mismatches 331; Indels 220; Gaps 43;

Qy 1 KLGIAEESKFKNU-----GNGKEGSILKMDITGVEHHHOENBESIKEKKSSFTIDRNISTI 54
Db 49 KLIQFNEIERSNLKVTSFDELKASSLK-----IDGIKTEMVIRENDKTRERNDFV 104.

Qy 55 R--DFENDDLKLKIKKF--REYDDFTE-----TGRMEYDYKDYDKGKN 96
Db 105 KFSEVENKMKLSSSELEVFKRKLDDLTETKETQSNOQRTLKILLERKEIE-----156

Qy 97 ITAYDDGTLEYETEKLD----BIKSKEYGVLSPSKDGHFELGKLSNVSNAKYVYGN 152
Db 157 -LVRVNNRNSNECKLRLSTIMPLETKQCGYTNDANSRTELERKTQELT----LLQSN 211

Qy 153 ---YKSTIBIKATK-Y-DPIHSKT--MTFDLYANINDIVCGLAFAGDMRLFVKND---QKKA 203
Db 212 DWLEKELRSKNEOYLTSYRQKTDVKWLDRLNRL----RNDFQMRNTNDVLKORN 265

Qy 204 EIKIRMPKEKE-----TKSEY PYVSSYYGVTELGEGDLS--KNPEDNLTKMS 250
Db 266 ELSKSLOQKLLIEKGLSPLSLNSEKQEEFAEMLSKQLRVLWDLLESQNLAKKEELNSIRELT 325

Qy 251 GKIYSDSKRQQYLKDNTILRKGYALKVTTYNGKTMDELGNSVSKREDIAKQKA---306
Db 326 AKVIADDKKQ-----TPENEDIILKELQL-TKEKLAQCEKECLRL 364

Qy 307 -----NPNLRALSETTIYADSRNVEDGRSTOSTLMSALDGFNTIYQVTFKMD 356
Db 365 SSITDEADNENLSAKSSSDFTFLKROLIKERTRKELQNBVPLINS 423

Qy 357 KGBAIDKGNLVTDSSKVLPGDKDKEYTGDKFVN2IKEGDSMLPTDTKPVNLSDKN 416
Db 424 FKERTDMLENLNAAUALL-----EITSNEK-NAK-VKE-----LNAKNQKLVCECN 468

Qy 417 YFPNSKSNKLYVNPENPEFLRGKLSDKGGFNWEILRVNE-SVVDNYLYVGDHLHDNTRDNI 475
Db 469 DLQTLTKQRDLICRQIQLLITNSVNSNSKGPLRKKEEOFQNMQEDDSTITESDSQKV 528

Qy 476 KJNVKDGDIIDWGMKDYKANGFDPKUTMDGNTYLTQTYSDLNAKAVYHOFQFLDNWK 535
Db 529 -----VTERLVEFKNIQLQ----EKAELUKV-VRNLAJKES 562

Qy 536 EVNIDPKGNTSIEYADGGSV-----VFNIN-----DKRNNFGDGEIQEHQHYINGKEY 583
Db 563 K---EKKRSQSLQRIESSTVNEAKEAITLTKSEKMDLIESRIEELKTSVPNEDA 619

Qy 584 TSPN-DIKQIIDKTLNIKIVVVD-----FARNNTYKEFILNKDTGEVSELKPHRYVTVT 635
Db 620 SYNVNTKQQLTKEKRDLESQYQDLQTISOTRESTENMSLINK--EIQDIXYSKDIS 676

Qy 636 IONGKEMSTTIVSEDF----ILPVYKGE--LERYQYQFDGEWEISREGKKA-----681
Db 677 IKLGKEKSSRILAERFLLNTLDTKAENDQLRKFEDY---LQNTLKLQDSKTHETL 732

Qy 682 -GYVINLISKDTFIPVFKIEEF-----KEEENKPTFDVSKKKDNPOVNHHSOLN 729

Db 733 NEVYSCKSKLSLTETELLNLKEEQKURVHLERNLQKBEINK---LSPKFDLSRIMVYQLO 788

Qy 730 ESHR-KEDL--QREHNSQKSSTDKVTA----TVLDKNNISSKSSTNNNP 772

Db 789 TLQEREDILEFTRSQCKKIDEBEALSELKETSQDKDHHKQLEEDNNNSN 840

RESULT 13
SLPM_BACBR
ID SLPM_BACBR
AC P06546;
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 01-FEB-2003 (Rel. 41, Last annotation update)
DB Middle cell wall protein precursor (MWP).
OS Bacillus brevis (Brevibacillus brevis).
OC Bacteria: Firmicutes; Bactillales; Peenibacillaceae; Brevibacillus.
RN NCBITaxonID=1393;
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN=47;
RX MEDLINE=88115203; PubMed=2828336;
RA Tsuboi A.; Uchibi R.; Adachi T.; Sasaki T.; Yamagata H.,
RA Tsukagoshi N.; Ueda S.;
RT "Characterization of the genes for the hexagonally arranged surface
RT layer proteins in protein-producing Bacillus brevis 47: complete
RT nucleotide sequence of the middle wall protein gene.";
RL J. Bacteriol. 170:935-945 (1988).
RN [2]
RN SEQUENCE OF 1-199 FROM N.A.
RC STRAIN=47;
RX MEDLINE=87137282; PubMed=3029027;
RA Yamagata H.; Tsuboi A.; Takao M.; Sasaki T.,
RA Tsukagoshi N.; Ueda S.;
RT "Cloning and characterization of the 5' region of the cell wall
RT protein gene operon in Bacillus brevis 47.";
RL J. Bacteriol. 169:1239-1245 (1987).
RN [3]
RN SEQUENCE OF 676-1053 FROM N.A.
RC STRAIN=47;
RX MEDLINE=0008404; PubMed=2428810;
RA Tsuboi A.; Uchibi R.; Tabata R.; Takahashi Y.; Hashiba H.; Sasaki T.,
RA Yamagata H.; Tsukagoshi N.; Ueda S.;
RT "Characterization of the genes coding for two major cell wall
RT proteins from protein-producing Bacillus brevis 47: complete
RT nucleotide sequence of the outer wall protein gene.";
RL J. Bacteriol. 168:365-373 (1986).
RN [4]
RN SEQUENCE OF 1-50 FROM N.A.
RC STRAIN=47;
RX MEDLINE=9-0078123; PubMed=2512285;
RA Tsuboi A.; Uchibi R.; Engelhardt H.; Hattori H.; Shimizu S.,
RA Tsukagoshi N.; Ueda S.;
RT "In vitro reconstitution of a hexagonal array with a surface layer
RT protein synthesized by Bacillus subtilis harboring the surface layer
RT protein gene from Bacillus brevis 47.";
RL J. Bacteriol. 171:6747-6752 (1989).
CC -1- FUNCTION: THE MIDDLE CELL WALL PROTEIN BINDS TO PEPTIDOGLYCAN AND TO
CC THE OUTER CELL WALL PROTEIN.
CC -1- SUBUNIT: THE MIDDLE CELL WALL LAYER IS COMPOSED OF SUBUNITS OF
CC THE MIDDLE CELL WALL PROTEIN. THESE PROTEINS FORM A HEXAGONAL
CC ARRAY WITH A LATTICE CONSTANT OF 14.5 NM IN THE MIDDLE CELL
CC WALL LAYERS.

CC	- - SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A HEPTAGONAL S-LAYER	Db	747 1PVEGDLTGAHYGQVIFGRKGGKTDIRWEKDSDDDLKE 803
CC	- - SIMILARITY: Contains 3 S-layer homology (SLH) domains.	Qy	505 DGVYLVGTGSIDLNAKAVGPHYQFLGVNVKPEVNIDPKGNTSEYADCKSVVNFINDGRN 564
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institutes of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).	Db	804 D-----IRNDEPIAFTVDNSBDDVVENVNNRKAQGMLA-BVTDERGMKDANIDGRN 856
CC		Qy	565 NGFDGEIOEQHQIYINGKEYTSFNDIKQIIDKTLINKIKIVVKDFARNTTYKEFLINKDTGEV 624
CC		Db	857 VGLVSDYRKDTI-----KASIKSATV-YFDLYDDFSEA 898
CC		Qy	625 SELKPHRVTVTQNKGEMSSTVSEEDFILPVKGEGLEKGYQFDGWETSGFGCKDAGYV 684
CC		Db	899 DG-----VNEGYVVMDSGDI-SGTRKD 923
DR	EMBL; M14238; AAP87321.1; -.	Qy	685 INFSKDPFIKPVFKIEP-----KCEBEENK-----TFDVSRRKDNPQVNHs 726
DR	EMBL; M19115; AAP22372.1; -.	Db	924 LIVSDAKTVRK-DKLEDDAEAFLKQPESERKDPPDKWALPSKEVGKFTSAGPKLYRA 981
DR	PIR; A28555; A28555.	Qy	727 -QLNESHKRKD1QREB---HSOKSDST---KDVTATV----LDKNNISSKSTNNPNK 773
DR	InterPro; IPR01119; SLH.	Db	982 TVELNSKVKAEDVDAFVFNGRKVEPSLLNFNPKDGVTIGTNTEDKVTSKIKVTNRNGK 1041
DR	Pfam; PF00395; SLH; 2.		
DR	PROSITE; PS01072; SLH DOMAIN; 2.		
KW	Cell wall; S-layer; SLH; Signal; Repeat.		
FT	SIGNAL 1 1053 AA; 117146 MW; DB4213188D95F4F CRC64;		
FT	CHAIN 24 1053 MIDDLE CELL WALL PROTEIN.		
FT	DOMAIN 29 92 SLH 1.		
FT	DOMAIN 93 143 SLH 2.		
FT	DOMAIN 144 203 SLH 3.		
SEQUENCE	1053 AA; 117146 MW; DB4213188D95F4F CRC64;		
SQ			
		RESULT 14	
		WAPA_BACSU	
		ID_WAPA_BACSU	
		STANDARD;	
		PRT;	2334 AA.
		AC Q07833;	
		DT 01-OCT-1994 (Rel. 30, Created)	
		DT 01-OCT-1994 (Rel. 30, Last sequence update)	
		DT 28-FEB-2003 (Rel. 41, Last annotation update)	
		DE Wall-associated protein precursor.	
		GN WAPA OR N17G.	
		OS Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.	
		OC	
		NCBI_TaxID=1423;	
		RN RP	
		SEQUENCE FROM N.A.	
		RC STRAIN168;	
		RX MEDLINE=93102506; PubMed=3116032;	
		RA SEQUENCE FROM N.A.	
		RC STRAIN=168 / BGSC1A1;	
		RX MEDLINE=95219088; PubMed=7704263;	
		RA Yoshiida K.-I., Shindo K., Sano H., Seki S., Oda M., Fujimura M., Fujita Y.; "Molecular analysis of three major wall-associated proteins of <i>Bacillus subtilis</i> subtilis 168: evidence for processing of the product of a gene encoding a 258 kDa precursor two-domain ligand-binding protein."	
		RT Cloning and sequencing of a 29 kb region of the <i>Bacillus subtilis</i> subtilis 168 genome containing the hut and wpa loci.",	
		RT Microbiology 141:337-343 (1995).	
		RL RN	
		RL Microbiol. 8:299-310 (1993).	
		RL RN	
		RP SEQUENCE FROM N.A.	
		RC STRAIN=168 / BGSC1A1;	
		RX MEDLINE=97124196; PubMed=8169500;	
		RA Yoshiida K.-I., Shindo K., Sano H., Seki S., Fujimura M., Yanai N., Miw Y., Fujita Y.; "Cloning and sequencing of a 65 kb region of the <i>Bacillus subtilis</i> genome containing the hut and wpa loci.",	
		RA Bacteriol. 142:3113-3123 (1996).	
		RC RN	
		RP SEQUENCE FROM N.A.	
		RC STRAIN=168 / BGSC1A1;	
		RX MEDLINE=98044033; PubMed=9384377;	
		RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G., Borchert S., Bertero M.G., Bessieres P., Bolotin A., Braun M., Brigneau S.C., Bron S., Borrius R., Bourrier L., Bruschi C.V., Caldwell B., Capuano V., Carter N.M., Brouillet S., Bruschi C.V., Connerton I.F., Cummings N.J., Daniel R.A., Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,	

Qy	566	--GFPGTQEHIYINGREYTSFNDIKOIDKTLNIKIVVDFARNTTVKEFLINKDTG	622
Db	1739	VSLSYGTDRVKSKSYNGTEKYIFT-----YDKNQNETSVVNK-EQNTKKRTPDNK---	1789
Qy	623	EYSELKPHRTVTIIONGKEMSSSTIVSEED-----FILPVYKG-----E	660
Db	1790	-----NRLTELTDRGSSQTWYPSDDKLTKTFWHDQKGTNQFTYNKLDDOMIMKD	1842
Qy	661	LEKGYQFDGWE----ISGPFGKDKDAGY-VINLSKDTFI----KPVFKK1BEEKEE	707
Db	1843	STSSYSPDYDENGVNQTFITGNGGTSFSYDERNLVSSLHIGDKNQGDILTESYEVDANG	1902
Qy	708	NKPTFDVSKKKNDPQNQHNSQLNEHSRKBDLQREPHSQ-----KSD	747
Db	1903	NRTTIN-SSASGRKVQYEKGKLNQ-----LVKETEDGTIVETTYDGFGNRKTVTIKDG	1955
Qy	748	STKDVDTATV-----LDKNNISSKSTNNNPNK	773
Db	1956	SSKTVNASFNIMNQLTKVNDES1SYDRNGNR	1986
RESULT 15			
CNA	STAAU	STANDARD;	PRT;
ID	CNA_STAAU		1183 AA.
AC	Q53655;		
DT	15-DEC-1998 (Rel. 37, Created)		
DT	15-DEC-1998 (Rel. 37, Last sequence update)		
DT	15-SEP-2003 (Rel. 42, Last annotation update)		
DE	Collagen adhesin precursor.		
CNA			
GN			
OS	Staphylococcus aureus.		
OC	Bacteria; Firmicutes; Bacillales; Staphylococcaceae.		
OX			
NCBI_TaxID	1280;		
RN			
RP	SEQUENCE FROM N.A.		
RC	STRAIN-FDA 574;		
RX	MEDLINE=92165839; PubMed=1311320;		
RA	Patti J.M., Jonsson H., Guss B., Switalski L.M., Wiberg K.,		
RA	Lindberg M., Hooeck M.,		
RA	"Molecular characterization and expression of a gene encoding a		
RT	Staphylococcus aureus collagen adhesin.";		
RT	J. Biol. Chem. 267:4766-4772(1992).		
RL			
RN	[1]		
RP	ERRATUM.		
RA	Patti J.M., Jonsson H., Guss B., Switalski L.M., Wiberg K.,		
RA	Lindberg M., Hooeck M.,		
RA	"Identification and biochemical characterization of the ligand		
RA	binding domain of the collagen adhesin from Staphylococcus aureus."		
RT	J. Biol. Chem. 269:11672-11672(1994).		
RL			
RN	[3]		
RP	COLLAGEN-BINDING DOMAIN.		
RC	STRAN=FDA 574;		
RX	MEDLINE=94032261; PubMed=8218209;		
RA	Patti J.M., Boles J.O., Hooeck M.,		
RA	Moore D., Jin L., Schneider A., Delucas L.J., Hooeck M.,		
RA	Narayana S.V.L.,		
RA	"Structure of the collagen-binding domain from a Staphylococcus		
RT	aureus adhesin";		
RT	Nat. Struct. Biol. 4:833-838(1997).		
RT	"Structure of the collagen-binding domain from a Staphylococcus		
RT	aureus adhesin";		
RA	J. Biol. Chem. 269:11428-11435(1993).		
RN	[4]		
X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS) OF 169-318.			
RP	X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS) OF 169-318.		
RX	MEDLINE=9475225; PubMed=934749;		
RA	Symvers J., Patti J.M., Carson M., House-Pompeo K., Teale M.,		
RA	Moore D., Jin L., Schneider A., Delucas L.J., Hooeck M.,		
RA	Narayana S.V.L.,		
RA	"Structure of the collagen-binding domain from a Staphylococcus		
RT	aureus adhesin";		
RT	Nat. Struct. Biol. 4:833-838(1997).		
RT	"Structure of the collagen-binding domain from a Staphylococcus		
RT	aureus adhesin";		
RA	J. Biol. Chem. 269:11428-11435(1993).		
RN	[5]		
CC	COLLAGEN-BINDING DOMAIN.		
CC	COLLAGEN-CONTAINING SUBSTRATA.		
CC	SUBCELLULAR LOCATION: Attached to the cell wall peptidoglycan by		
CC	an amide bond (Potential).		
CC			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -		
CC	the European Bioinformatics Institute. There are no restrictions on its		
CC	use by non-profit institutions as long as its content is in no way		

modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

```

DR EMBL; M81736; AAA20874.1; -
DR PDB; 1AMX; 24-JUN-98.
DR PDB; 1D2O; 27-SEP-00.
DR PDB; 1D2P; 27-SEP-00.
DR InterPro; IPR001899; Gram_pos_anchor.
DR InterPro; IPR06192; LPXTG.
DR TIGRFAMS; TIGR01167; LPXTG anchor; 1.
DR PROSITE; PS00847; Gram_pos_ANCHORING; FALSE_NEG;
KW Cell wall; Peptidoglycan_anchor; Repeat; Signal; 3D-structure.
FT SIGNAL 1 29 .
FT CHAIN 30 1154 .
FT PROPEP 1155 1183 .
FT DOMAIN 151 318 .
FT DOMAIN 533 1093 .
FT DOMAIN 1093 1157 .
FT REPEAT 533 719 .
FT REPEAT 720 906 .
FT REPEAT 907 1093 .
FT SITE 1151 1155 .
FT MOD_RES 1154 1154 .
FT STRAND 174 179 .
FT TURN 182 183 .
FT TURN 185 186 .
FT STRAND 187 194 .
FT TURN 196 197 .
FT STRAND 201 201 .
FT STRAND 205 211 .
FT STRAND 215 228 .
FT TURN 229 230 .
FT STRAND 232 234 .
FT HELIX 239 246 .
FT TURN 248 249 .
FT STRAND 251 255 .
FT TURN 256 259 .
FT STRAND 260 265 .
FT HELIX 267 270 .
FT TURN 271 272 .
FT STRAND 273 283 .
FT TURN 286 287 .
FT STRAND 290 299 .
FT STRAND 301 301 .
FT TURN 302 303 .
FT STRAND 307 311 .
FT STRAND 314 317 .
SQ SEQUENCE 1183 AA; 133066 MW; B611CC072E575D76 CRC64;

Query Match 4.5%; Score 179.5; DB 1; Length 1183;
Best Local Similarity 18.7%; Pred. No. 0.77; Matches 172; Conservative 128; Mismatches 325; Indels 297; Gaps 42;
Matches 172; Mimatches 325; Indels 297; Gaps 42;

Qy 22 KKDITGTYBHHBQNEESIKEKSFTIDRNISTIRDENKDLK-----KLIIKKFREV 73
Qy 304 KEEVNGKSFNHTVHNINANAGLEGTIVKGELVKL-QDKDTKAPIANVKFLSKKGDSVV 361
Db 74 DDFTSSETGKRMEEYDYKDDKG---NTIAYDGTDLYEETE-----KLDEIKTSKIVYLSP 126
Db 362 KD-----NOKEIEIITDANGIANKALPSGDYIYKEIAPRPTFDKDCEYPMKDT 414
Db 415 DNGGYFTTENAKAIKETKDVSAQKWTGTO----KVKTPTY-----452
Db 182 VDGGLAFAGDMRLFEKNDQK----KAEKIKRMPKEIKETKSEYPVSSGNVIEL---232
Db 453 -----FVKYKODDNQNTTPVDAEIKKLEDGTTKVWSNLPENDNGKAIVLVKE 503
Qy 233 -GEGDLSKNEPDNLTKMESEKYSIISDSEKQOYLLKNIILRGALKVTTNPGBTMDLB 290
Db 504 VNAQGE-----DTPBEGTYKKGKGLVNTNEK-----PIETTSIS 538

```

Qy 291 GNGVYSKEDIAKIQKANPNRALSETTYADSRNVEDGRSTOSVLMSALDGENITRYQVF 350
 Db 539 GEKW-----DDKDNDQGKRPEKSVNLANGEKVFLDV 573
 Qy 351 TFKMDKGEAIDKDGNLWTDSSKLVLFGKDDKEYGEDKFENVAIKEGSMMLFIDTKPVN 410
 Db 574 TSETNWKYEF--KDLPKYDEGKCI-----EYT-----VTEDHVKDY--ITDIN 612
 Qy 411 LSMDKNYNPNSKSNKIYRN-----PEFYLRGKISDGFF----NWE- 448
 Db 613 GTITNKTPGETSATVTKWDDNNNQGKRPETEIKVELYDQGKATGKTAILNESNNWTH 672
 Qy 449 -----LRYNESTVDNYLIYG-DLHIDNTRDFNKLNVK-----DGDIMDWG 488
 Db 673 TWGLDEAKGQOQVKYTBEILTKVKGYTHVNDMGRILVNKYTPETTSISGEKV-WD 731
 Qy 489 MKDYKANGFPDKYU-----DMDGNYVLQGYSIDLNAKAVG--VHYQFLYDNV 533
 Db 732 DKDNQDGRPEKSVNLJADGEVKTLJDTSETNWKYEFKDLPKYDEGKCIETYTEDHV 791
 Qy 534 KPEVNIDPKGNT-SIEYAGK--SVVENINDKRNNGEDG---EIQOQHIVINGK--- 581
 Db 792 K-DYTTDINGTITNKYPGETSATVTKWDD--NNNGDGRPETEIKE-LYQDGKATGK 847
 Qy 582 -----EYTSFNDIKOIDIKTLNIKVVKDFARNTVKEFLNKDGE 623
 Db 848 TALLNESSNNWHTWTGLDEAKGQOQVKYTVEELTKVGYTHVNDNNDGNLIVTNKYPE 907
 Qy 624 VSEL-----KPHRVTVT IONGKEMSTTIVSEEDEFILPVYKGELENGXQF 667
 Db 908 TTSLSGEKWDDKDNDQGKRPEKSVNLANGEVKTLJDTVS-----TNWKEFVKDLPKY 963
 Qy 668 DGWEISGEGK-----DAGYVINLSKDFPIK-----PVFKLLEKKBEPBNK 709
 Db 964 D-----EGKCLLEYTVTEDHVKDYTTINGTITNKYPGETSATVTKWDDNNNQGK 1016
 Qy 710 -PT-----FDVSKKKDNPQVNHSQLNESHRKEDLOREHSDSTKDVTATV----- 756
 Db 1017 RPEIKVELYQDGKATGKTAILNESNNWHTWTGLDEAKGQOQVKYTDELTKVNGYTH 1076
 Qy 757 LDKN-----ISSKSTTNNNPNK 773
 Db 1077 VDNNDMGNLIVTNKYPCKPNK 1098

Search completed: January 28, 2004, 13:00:44
 Job time : 24 secs

Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 28, 2004, 12:57:43 ; Search time 20 Seconds (without alignments)

Title: US-10-067-385-8

Perfect score: 4026

Sequence: 1 KLGRIAESKFKNLNGKEGS.....ATVLDKNNISSKTTNNPNK 773

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing First 45 summaries

Database : PIR_76:*

1: Pir1:*

2: Pir2:*

3: Pir3:*

4: Pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query	Match	Length	DB ID	Description
1	4026	100.0	2140	2	F95074	serine proteinase,
2	3709.5	92.1	2144	2	A97942	metalloproteinase
3	253.5	6.3	1127	2	T18317	ORF MSV156 hypothe-
4	250	6.2	1650	2	T1844	hypothetical prote
5	235.5	5.8	3724	2	T18427	hypothetical prote
6	231.5	5.8	2269	2	T28677	rhoptry protein -
7	230	5.7	2401	2	T28676	rhoptry protein -
8	228	5.7	4550	2	T18440	hypothetical prote
9	221.5	5.5	1711	2	T18429	probable major sur-
10	219	5.4	1640	2	A24594	major merozoite su-
11	218	5.4	1631	1	SA2601	lipoprotein [impor-
12	218	5.4	1639	2	S02603	reticulocyte-bindi-
13	217.5	5.4	1546	2	G90603	DNA-directed RNA p-
14	215.5	5.4	1252	2	B42771	membrane nuclelease,
15	214	5.3	2339	2	A45597	hypothetical prote
16	212.5	5.3	1125	2	E90598	chromosome Segrega-
17	212.5	5.3	2166	2	G70163	hypothetical prote
18	212.5	5.3	2485	1	H71621	probable cell surf-
19	212	5.3	1188	2	A71621	mature-parasite in
20	211.5	5.3	1104	1	A36866	RAD2 endonuclease
21	209	5.2	1558	2	B71603	hypothetical prote
22	207.5	5.2	4688	2	F82885	chromosome Segrega-
23	207	5.1	1141	2	E89824	hypothetical prote
24	207	5.1	1191	2	B97116	hypothetical prote
25	207	5.1	1979	2	C71622	probable cell surf-
26	206.5	5.1	1622	2	AE7171	mature-parasite in
27	206	5.1	1526	2	A45605	RAD2 endonuclease
28	204.5	5.1	1516	2	E71619	hypothetical prote
29	204.5	5.1	2500	2	G71609	impl protein - Myc

GAF domain protein
hypothetical prote
fibrinogen-binding
hypothetical prote
superfamily I DNA
repeat organellar
hypothetical prote
AMP-dept; acyl-CoA
hypothetical prote
surface Protein (L)
hypothetical prote
Ser/Thr protein ki
hypothetical prote
hypothetical prote
surface-located me

Impl protein - Myc

ALIGNMENTS

RESULT 1

F95074 serine proteinase, subtilase family [imported] - Streptococcus pneumoniae (strain TIGR4)

C;Species: Streptococcus pneumoniae

C;Date: 03-Aug-2001 #sequence_change 03-Aug-2001

C;Accession: F95074

R;Tettelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Heilon, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapple, T.; Hickey, E.K.; Holt, I.E.

Science 293, 498-506, 2001

A;Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison, A;Title: Complete Genome Sequence of a Virulent Isolate of Streptococcus pneumoniae.

A;Reference number: A95000; PMID: 11463916

A;Accession: F95074

A;Status: preliminary

A;Molecule type: DNA

A;Cross-references: GB:AB005672; PIDN:AAK74791.1; PID:914972117; GSPDB:GN00164; TIGR:SP

A;Experimental source: strain TIGR4

C;Genetics:

A;Gene: SP0641

Query	Match	Best Local Similarity	Score 4026;	DB 2;	Length 2140;
Matches	Conservative	100.0%; Pred. No. 4e-167;	100.0%; Pred. No. 4e-167;	Indels	Gaps 0;
Qy	1	KLGIAESKFKNLNGKEGS.....ATVLDKNNISSKTTNNPNK	60		
Db	1334	KLGIAESKFKNLNGKEGS.....ATVLDKNNISSKTTNNPNK	1393		
Qy	61	DLKLKLKKKFREVDFTSETGRMEEYDYGDDKGNTIAYDGDTIUYETEKLDEIKSKI	120		
Db	1394	DLKLKLKKKFREVDFTSETGRMEEYDYGDDKGNTIAYDGDTIUYETEKLDEIKSKI	1453		
Qy	121	YGVLSPKSDKGHFBILGKISNVSKNVAKVYGNINYKSTIKATKDYHFSKTMFTDLYANIND	180		
Db	1454	YGVLSPKSDKGHFBILGKISNVSKNVAKVYGNINYKSTIKATKDYHFSKTMFTDLYANIND	1513		
Qy	181	IVDGGLAFAGDNRLFLVTFDNDQKKAETKIRMPKETKETKSEPVYVSGNVIELGGDLSKVN	240		
Db	1514	IVDGGLAFAGDNRLFLVFDNDQKKAETKIRMPKETKSEPVYVSGNVIELGGDLSKVN	1573		
Qy	241	KPDNLTKMESGTYISDSEKQOYLKDONTILRGYALKVTTNPKGTDMLBENGVYSKEDI	300		
Db	1574	KPDNLTKMESGTYISDSEKQOYLKDONTILRGYALKVTTNPKGTDMLBENGVYSKEDI	1633		
Qy	301	AKIQKANPNLALSETTIYADSRYEDGRS7QSVLMSALDGFNITRYQVFTEKMDKGEA	360		
Db	1634	AKIQKANPNLALSETTIYADSRYEDGRS7QSVLMSALDGFNIIYQVFTEKMDKGEA	1693		
Qy	361	IDKDGNLVTDSSKLVLFGKGDKEYTEDKVNTEAKEDGSMLFIDTPKPVNLSMDQNYFNP	420		

Db	1634	IDKDGNLVTDSSKVLVLFGKDDKEYTEDKPNVEAIKEDEGSMFLIDTKPVNLMSMDKRYFNP	1753	Qy	301	AKIQKANPNIRALSETTIYADSRNVEDGRSTOSVILMSALDGPNIIRYQVFPMNDKGREA	360
Qy	421	SKSNKIVYRNPEFYLRGKISDKGGENWELRVNESWDNYLIVGDHLHDNTDRDNKLNK	480	Db	1638	AKIQKANPNLRVSETTIVADSRNVEDGRSTOSVILMSALDGPNIIRYQVFPMNDKGREA	1697
Db	1754	SKSNKIVYRNPEFYLRGKISDKGGENWELRVNESWDNYLIVGDHLHDNTDRDNKLNK	1813	Qy	361	IDKDGNLVTDSSRLVLFGKDDKEYTEDKPNVEAIKEDEGSMFLIDTKPVNLMSMDKRYFNP	420
Qy	481	DGDIMDWMKDYKANGFDKVTDMDGNYLQGYSUNAKAVGVRYOFLDVNKPEVNID	540	Db	1698	IDKDGNLVTDSSKLVLFGKDDKEYTEDKPNVEAIKEDEGSMFLIDTKPVNLMSMDKRYFNP	1757
Db	1814	DGDIMDWMKDYKANGFDKVTDMDGNYLQGYSUNAKAVGVRYOFLDVNKPEVNID	1873	Qy	421	SKSNKIVYRNPEFYLRGKISDKGGENWELRVNESWDNYLIVGDHLHDNTDRDNKLNK	480
Qy	541	PKGNTSIEYADGKSVVFNINDKRKNNGFGEIOEHQHITYNGKEYTSFNDIKQDCKTLNK	600	Db	1758	SKSNKIVYRNPEFYLRGKISDKGGENWELRVNESWDNYLIVGDHLHDNTDRDNKLNK	1817
Db	1874	PKGNTSIEYADGKSVVFNINDKRKNNGFGEIOEHQHITYNGKEYTSFNDIKQDCKTLNK	1933	Qy	481	DGDIMDWMKDYKANGFDKVTDMDGNYLQGYSUNAKAVGVRYOFLDVNKPEVNID	540
Qy	601	IIVKDFARNTTYKEFLINKDGEUSELKPHRVTVTQNGKENSSTIUSSEDFLPVYKGE	660	Db	1818	DGDIMDWMKDYKANGFDKVTDMDGNYLQGYSUNAKAVGVRYOFLDVNKPEVNID	1877
Db	1934	IIVKDFARNTTYKEFLINKDGEUSELKPHRVTVTQNGKENSSTIUSSEDFLPVYKGE	1993	Qy	541	PKGNTSIEYADGKSVFNINDKRKNNGFGEIOEHQHITYNGKEYTSFNDIKQDCKTLNK	600
Qy	661	LEKGYQFDGWESFGFKRDKAGYVNLKDTFLKPVKPKIEKEEKKPDKVSKKDN	720	Db	1878	PKGNTSIEYADGKSVFNINDKRKNNGFGEIOEHQHITYNGKEYTSFNDIKQDCKTLNK	1937
Db	1994	LEKGYQFDGWESFGFKRDKAGYVNLKDTFLKPVKPKIEKEEKKPDKVSKKDN	2053	Qy	601	IIVKDFARNTTYKEFLINKDGEUSELKPHRVTVTQNGKENSSTIUSSEDFLPVYKGE	660
Db	2054	PQVNHSQLNESHRKEDLQREHEQSQSDSTKDVATVLDRNNISSKSTNNPNK	773	Db	1938	IIVKDFARNTTYKEFLINKDGEUSELKPHRVTVTQNGKENSSTIUSSEDFLPVYKGE	1997
Qy	721	PQVNHSQLNESHRKEDLQREHEQSQSDSTKDVATVLDRNNISSKSTNNPNK	773	Qy	661	LEKGYQFDGWESFGFKRDKAGYVNLKDTFLKPVKPKIEKEEKKPDKVSKKDN	720
Db	2054	PQVNHSQLNESHRKEDLQREHEQSQSDSTKDVATVLDRNNISSKSTNNPNK	2106	Db	1998	LEKGYQFDGWESFGFEKSKDAGYVNLKDTFLKPVKPKIEKEEKKPDKVSKKDN	2057
<hr/>							
RESULT 2							
A97942	metalloproteinase (EC 3.4.21.-) A [imported] - streptococcus pneumoniae (strain R6)						
C;Species:	Streptococcus pneumoniae						
C;Date:	22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 02-Nov-2001						
C;Accession:	A97942						
R;Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszczak, L.; Burgett, S.; DeHoff, B.S.; e; R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.; Y.; P.; Sun, P.M.; Winkler, M.E.	ORF MSV156 hypothetical protein - Melanoplus sanguinipes entomopoxvirus						
J;Bacteriol. 183, 5709-5717, 2001	C;Species: Melanoplus sanguinipes entomopoxvirus						
A;Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.; A;Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.	C;Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 21-Jul-2000						
A;Reference number: A97942; PMID:21429245; MUID:211544234	C;Accession: T28317						
A;Status: Preliminary	R;Afonso, C.L.; Tulman, E.R.; Lu, Z.; Oma, E.; Kutish, G.F.; Rock, D.L.						
A;Molecule type: DNA	J;Virol. 73, 53-552, 1999						
A;Residues: 1-2144 <KDR>	A;Title: The genome of Melanoplus sanguinipes entomopoxvirus.						
A;Cross-references: GB:AE007317; PID:AAK99365.1; PMID:915458138; GSPDB:GN00174	A;Reference number: 220484; PMID:9812612; PMID:99102612						
A;Gene: prta	A;Accession: T28317						
C;Keywords: hydrolase; serine proteinase	A;Status: Preliminary; translated from GB/EMBL/DDBJ						
C;Genetics:	A;Cross-references: EMBL:AF063866; NID:94049647; PID:94049717						
C;Genetics:	A;Residues: 1-1127 <AF0>						
C;Genetics:	A;Molecule type: DNA						
C;Genetics:	A;Cross-references: EMBL:AF063866; NID:94049647; PID:94049717						
Query Match	92.1%	Score 3709.5;	DB 2;	Length 2144;			
Best Local Similarity	92.4%	Pred. No. 2,2e-153;	Indels 1;	Gaps 1;			
Matches 714; Conservative	26;	Mismatches 32;					
Qy	1	KLGIAESEKFKNLGNKGKGSLLKDDTGYEHHQENEESIKEKSFTIDRNISTIRDENK	60	Db	1339	KLGIAESEDKLNKASPEINT-NNNQAKEDSKPDEKSVGEASELSNEINTKTIREPENK	1397
Db	61	DLKLKLIKFFREVDFTSETGRMEEYDKGKNTIAYDGTDLVETEKLDEIKSKI	120	Qy	61	KLGLKLIKFFREVDFTSETGRMEEYDKGKNTIAYDGTDLVETEKLDEIKSKI	1517
Db	1398	DLKLKLIKFFREVDFTSETGRKTEEYDKDGNNTIAYDGSALQYETEKFDEIKSKI	1457	Db	1398	YVGLSPLSPSKDGHFEILGKLSNVSKNAKYGGNNYKSIEIKATYDHSHSTMIFLYANIND	240
Qy	121	YVGLSPLSPSKDGHFEILGKLSNVSKNAKYGGNNYKSIEIKATYDHSHSTMIFLYANIND	180	Db	1518	YVGLSPLSPSKDGHFEILGKLSNVSKNAKYGGNNYKSIEIKATYDHSHSTMIFLYANIND	1577
Db	1458	YVGLSPLSPSKDGHFEILGKLSNVSKNAKYGGNNYKSIEIKATYDHSHSTMIFLYANIND	240	Qy	181	YVGLSPLSPSKDGHFEILGKLSNVSKNAKYGGNNYKSIEIKATYDHSHSTMIFLYANIND	300
Qy	241	KPDNLTKMEKSGKTYSDSEKQYLKDNTLRLGYALKTTYNQGKTMLEGNGCVYSKEDI	1637	Db	1578	KPDNLTKMEKSGKTYSDSEKQYLKDNTLRLGYALKTTYNQGKTMLEGNGCVYSKEDI	269
Db	1638	AKIQKANPNLRVSETTIVADSRNVEDGRSTOSVILMSALDGPNIIRYQVFPMNDKGREA	1697	Qy	217	--KSEY--PYVSSYCNVIEEGDLSKRNPDNLTKMEKSGKTYSDSEKQYLKDNTLRLGYALKTTYNQGKTMLEGNGCVYSKEDI	2425
Db	1698	IDKDGNLVTDSSRLVLFGKDDKEYTEDKPNVEAIKEDEGSMFLIDTKPVNLMSMDKRYFNP	1757	Db	2116	LFDNDIQKLNNIDTFEYKIIDKNNNQLKEELEYSK	269
Db	1758	AKIQKANPNLRVSETTIVADSRNVEDGRSTOSVILMSALDGPNIIRYQVFPMNDKGREA	1817	Db	2116	EFYKIDKNNIDTFEYKIIDKNNNQLKEELEYSK	269
Db	1818	DGDIMDWMKDYKANGFDKVTDMDGNYLQGYSUNAKAVGVRYOFLDVNKPEVNID	1877	Db	2116	EFYKIDKNNIDTFEYKIIDKNNNQLKEELEYSK	269
Db	1878	DGDIMDWMKDYKANGFDKVTDMDGNYLQGYSUNAKAVGVRYOFLDVNKPEVNID	1937	Db	2116	EFYKIDKNNIDTFEYKIIDKNNNQLKEELEYSK	269
Db	1938	IIVKDFARNTTYKEFLINKDGEUSELKPHRVTVTQNGKENSSTIUSSEDFLPVYKGE	1997	Db	2116	EFYKIDKNNIDTFEYKIIDKNNNQLKEELEYSK	269
Db	1998	LEKGYQFDGWESFGFKRDKAGYVNLKDTFLKPVKPKIEKEEKKPDKVSKKDN	2057	Db	2116	EFYKIDKNNIDTFEYKIIDKNNNQLKEELEYSK	269
Db	2058	PQVNHSQLNESHRKEDLQREHEQSQSDSTKDVATVLDRNNISSKSTNNPNK	2110	Db	2116	EFYKIDKNNIDTFEYKIIDKNNNQLKEELEYSK	269

A; Molecule type: DNA
A; Residues: 1-3724
A; Cross-references: EMBL:Z98547; NID:e1325376; PID:e1325379; PID:CAB11104.1
C; Genetics:
A; Introns: 307/1; 1545/2
A; Note: C0325c

Query Match Score 235.5; DB 2; Length 3724;
Best Local Similarity 18.7%; Pred. No. 0.032;
Matches 192; Conservative 150; Mismatches 312; Indels 371; Gaps 45;

5 IAPSKFKLNGKEGS-LKK-----EKSSPT 46
5 IAPNLFENILNNQHNDLQRNITLMDVNTDPLEHQEONQLNKNKCLGTGKCKYM 308

47 I-----DRNISTRDFENKDLRKLRKKFREVDFSETGRMEEVDYKDDKG 95
309 IPKGTEYQDKERKESILTINQNDKKYKKK-KKKYSELODSNISSNTLSKRYTT--- 363

96 NIAYDDSTDLLEYETEKLDEIKSKIYGVLSPSKDGHFFILGKISNVSKNAKYGGNNKS 155
364 -----CGMDKETKEDQEQR-----RNTEAVICVNVCNVTNEKORKKCTNKKK 407

156 IEIKATKDEHSKTMTPFLYANINDIVDGLAFGDMR-LFVKNDQKAEKIRKMP-EK 212
408 --SVQKDVLVEENALLD--NU-IIIGINFDDVTRKECKLIDDDNNNNENIDNNIYVENK 460

213 IKETKSEPYVSSYG-NVI-----ELGEGDLSKPNPDLTKMESGK 252
461 LRDQKS-YDLSSEGSKNLLGYNEGEEFNEEVENIEKELEKQDKEKKNDK---T 512

253 IYSDSEKQOYLKDNTILRKVALKVTYNPGBTIDMLEGNGYYSKEDAKIOKAN-PNLR 311
513 IYNNNEBDTDLNRNI-----NKIEB1NNNND--NNNNNNKKENKARTEHILNKE 562

312 ALSETTIYADSRNVEDGRSTQSULMSALDGFDNIRYQWTFKNDKGEAIDKDGNLVTS 371
563 SISKHKGSPSRDK-----IKLYTTKNEDESTFELKELEITNN 604

372 SKULVFGKDKEYTDKFNVAEAIKEDSMSLFDTKPKVNLSDKGNPNSKSN----- 424
605 KVNVY-BEEDIIGSNEDEDYIHLKE-----NLKEDANEYNDKENKNNKTK 650

425 -----KIVYR-----PEFYLRGKLSDKGGFWMLRVN--ESVVDNYLIYGDHLHIDNTRD 472
651 ILKSKNYLENERKTLEELKLRKG--NNIFKDEKTNLSLGEVINEIQNEENKINDIQ 707

473 ENK-----LNVKD--GJIMDWGMKDYKANGFPDKYTMDF----- 505
708 GNISKQKLIQSSRTNDPENKIDSLNNDLEKEKRKCSQHFIDNLYKADKNEISENTNK 767

506 -----GVNLQT-----GYSDLNKAQGVHY----- 526
768 ICDDNNNNYIDESNNYDBSINNYDENINNNYDENINNNYDENINNNYDENINNNYDE 827

527 -----QFLYDNYKPEVNID-----PKGNTSIE-----Y 549
708 KENYDIFINKIKRESLLKIDKN1KDNNDNDEYIMDNFYENDFTLHNKHEITNKELDPL 947

550 ADGKSVVFN-----INDKRNNGFGEIOE----- 573
888 KENYDIFINKIKRESLLKIDKN1KDNNDNDEYIMDNFYENDFTLHNKHEITNKELDPL 947

574 -----QHIIYINGKEYTS--FNDIKQIIDIKTLNK1KVKDPAARTVKEFLINKDTGB 623
948 NTQNEBFIENIDIKKKYTNDHFENDADMFYEMN-KLNLKOMKKKQEFKFTDETG 1005

624 VSELK-----PHRTVUTONGKEMISSTVSE 649
1006 LQSHIKIKYNGKEKHDKQNNEEKNILLDENQVTSVLNSDHKEQD1IHSQTNCIDE 1065

650 EDFILPVYKGELEKGYOFDGWEISGFEGKDGAGTVINLISKDFT1KPVFKIEFKKEEEENK 709

Db 1066 NN-IEQINEENSKKGVRISSGTD--ENKND-----MENKNDMEKK 1102
Qy 710 PTFDVSKKDNQPOVNH-SQNLNESSHREKDFOREHHSQKSDSTKDVTATVLDKNNISSLKSTT 768
Db 1103 N-DMEKNDTEKNDMEKNDMEKNDMEKNDMEKNDMEKNDME-----NENNMENRNSDI 1153
Qy 769 NPNNK 773
Db 1154 ENENK 1158

RESULT 6
T28677
rhoptry protein - Plasmodium yoelii
C.Species: Plasmodium yoelii
C; Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 01-Dec-2000
C; Accession: T28677; C45521
R; Keen, J.; Siiba, K.; Holder, A.
Mol. Biochem. Parasitol. 65, 171-177, 1994
A; Title: A gene coding for a high molecular mass rhoptry protein of Plasmodium yoelii.
A; Reference number: Z20508; MUID:95021522; PMID:7335623
A; Accession: T28677
A; Status: preliminary; translated from GB/EMBL/DDBJ
A; Molecule type: DNA
A; Residues: 1-269 <KEE>
A; Cross-references: EMBL:127838; NID:9457145; PID:9457146; PIDN:AAA21304.1
R; Keen, J.; Holder, A.; Playfair, J.; Lockyer, M.; Lewis, A.
Mol. Biochem. Parasitol. 42, 241-246, 1990
A; Title: Identification of the gene for a Plasmodium yoelii rhoptry protein. Multiple c
A; Reference number: A45521; MUID:91101660; PMID:2270106
A; Accession: C45521
A; Status: preliminary
A; Molecule type: DNA
A; Residues: 2131-2269 <KE2>
A; Cross-references: GB:M34283

Query Match Score 231.5; DB 2; Length 2269;
Best Local Similarity 19.7%; Pred. No. 0.025;
Matches 181; Conservative 147; Mismatches 330; Indels 259; Gaps 377;

Qy 37 ESIREKSSFTIDRNISTIRDPEK-DLKLLKKFREVDDFTSETGKRMEEY----- 87
Db 109 EGKLELNKKIKDIAKLEYVKNTVELKKEIEKNNAVYDELANOSSPKVTKGTENKNTYI 168

88 -----DYYKDDKGNIAYDDGTDLEYETEKL----- 227
Db 169 NTIKSYFDQIYEGDIDPYNELOSSIVEKDPIDEDK-T-KLENLRSKIDNVDYDKIQKMEI 227
Qy 133 E-ILGKISNVSKNAKYGGNNKSIEIKATKDEHSKTMTPFLYANINDIVGLAFGDM 191
Db 228 EYTKSHANNIETNNKL---PNTLIEKTYIDEISK---BLNKMLEDFKNK---EKL 276

Qy 192 RLFDPKNDQKK--AEIKIRMP-----KIKETKSEPYVSSYGVNVILEG 234
Db 277 SNKSDYDKGKREOLESEYKSKMELIRNHYNQSNTNDKTEEAKONTYDKSNEHMTIPTNE 336

Qy 235 GDL-SK-----NKPDLNLTKME-----GKIYSD---SEKQQYL-LKDNTI--ILRKSYAL 276
Db 337 DEISKI1SEVKTMRKDEILSKVNTYIDPNKCKYKETVNSEHSQFTELDKIAEVSKELKK 396

Qy 277 KVTTYNFGKTDMLEGNGVYSE--DIAKIOKANPLRALSET-----TIVAD-- 321
Db 397 CBOFNDNSKLINETKNSIEQYQININTLKKCYDEYIKVCKSTKESTYKFSQTLKDM 456

Qy 322 SRNFGRSTQSVLMSALDGFNIRYQVFT-----FMN-----DKGEAI--- 361
Db 457 NQNIKTVKETNSIDSYSIEKFPE---QILTGKOTKLENKFTBSLNHEANNNELKIFYFS 512

362 DKDQNLVYDSSKLVLFGKDKEYTGEDKFSMLFDTKPVNLSDMKYFNPS 421
513 DLKANLG1NEENMLYNOFTEKEKTFND-----IKERNHINHEEISKIEKIHASTINIS 566

422 KSNKIIYVRNPEFVRGK1SDKGFFNWLVRVNEVVDNYLIYGDHIDTRDFNKLNVK- 480

Db	567 ETEREIGINIESLNTKPEK-----VKENVTNLKIKBKLKHYSDFRGKEGNKRY	618	Qy	152 NYK-IEBIKATYDFHSKTMFDLYAN-----INDI---VGGLAFAFGDMRLFVKDNDQKKA 203
Db	481 -----GCDIMDWGMK-DYKANGFPD-----KVTDMDGRV-----	520	Db	743 KIRNDDIDTLNQKIDSLSLETLT-EIKRGENSHIDEIKQIDKLUKVKPRTKME--NEDPKREI 799
Qy	619 TDKKKINNDINMVAISQDQHNGLDDQIKRSBSVYSEIKLERSNTBISNDNE 678		Qy	204 EIKI-RMPEKIKETKSEPYVSSYGNVIELGEGD--LSKRNKPDLNITKOMES-GKIV-----254
Db	521 AVGVHYQFLDYNVKPEYNNI-----DPKGNTSIEYADGKSVVENIN-----560		Db	800 ERKKEENIVEKLDKKKNTYKEIDKLKNEISKLEKLNINISYGKSLGNPLQOI 859
Qy	679 GIKKKOQI1TVIDKKKNTYEEINKLJSEISKIEKONTSLERKVKDINLSYGNLGNLFLE 738		Qy	255 -SDSEKOQYLK-----DNI-----ILRKGYA 275
Db	561 -DKRANGSFDEILOEQHAYIN-----GREYTSFNIDIQQLIKTLN-IKI-----601		Qy	860 DEEKCAELHTKAMEAYIDDDNKKKSQEIEKEMMNINMDIKMDIHKEMKALNISHDYK 919
Qy	739 QIDEEKKAENTIKSMRAYIDDDNKKKSQEILETENDIKDINKEMEALKISHDDKKC 798		Qy	276 LKVTT--YNPGKTDMILEGGVYSKEDIAKIQKANPNLRALESSTTYADSRAYEDGRSTQS 333
Db	602 -----VYKDFARNNTVKEFILNKDTGEVSELKLPHRVTUTQNGKE 641		Db	920 IYHTTSKKNHEEKISDIRKNSLKIQDSESESYINDIKELEKVNVLBESQNNTDINQ----975
Qy	799 HDKSKRNHKENISDIYDKSSKLQDFSRBEND-IDKRLQKNSSESNNHNSDIN-----851		Qy	334 VLMSLDG-FNIIKYQVFTFKMNDGEAIDKDGNYLYTDSKVLFLFGKDKEYTGE-----387
Db	642 MSSTIVSSEDFILPVYKGLELKGYQFDQWEISGPEGKDKAGYVINLSKDFTIKPVFKIE 701		Db	976 -YLSKTCIENIYNL-----KLANKKIKIIDK-----VKEYTDETEBKNN 1010
Qy	852 -----OCLINEVANIYNTLKLNLKIKLIDRKV 877		Qy	388 DKEFNV-----AIKEDGSMMLFLDTKPVNLSMDKKNYFNPSSKSNKLYVRNPEFYLRKG 438
Db	702 EKKEENNPPTFDVSKKDNPQVNHQSOLNEHSHRKEDLOREEHQSQSDST-----KDV-----T 753		Db	1011 KKINAELSNSEKITOLKENSSLKECOSK-1KSTDIDNNYSECIKN--ITNLKTYI---1063
Qy	878 EYTSIEK-----NGKNNDELNNSEKVIKLEGDSLSEKCRSKINSLDDKXIDECIKN 932		Qy	439 ISDGGEFNWELRVNESYVDNT-LIYGDLDHIDTRDFNLTQKNGDGDIMDWGMKDYKANGK 497
Db	754 ATYLDKNNNISKSSTNN 770		Db	1064 VNEKANNINTYFKNAEETYNQVNLNFNTIEMAITSQYI-LNPK-----KNG-1109
Qy	933 INVLKNNLNEETNITN 949		Qy	498 PDKVTDMDGNYVLTQGTYSIDLNAKAVGHYOFIYDNYKPEVNIDPKGNTSIEYADGKSVVF 557
Db	RESULT 7		Db	1110 -----TNTNTDNTYNTKELKEHKK-----KSNVYKDEAGRNTQEIKKONKEL--1147
T28676	rhoptry protein - Plasmodium yoelii (fragment)		Qy	558 NINDKRANGFDGEI1QEHIIYNGKEYTSFNDIKQIDKTLNI-KIVVKDF--ARNTYKE 614
C;Species: Plasmodium yoelii			Db	1148 -----FEKYEEQETVLLN-KYTAV--EUKNKEDKTKNYSQEIKEIKDAHNTFTSQ 1195
C;Accession: R26676; A45521			Qy	615 FILNKDGTGEVSLPHRVTVTIQNQGEMSTSIVSEEDFILPVYKQGKQYDGFQGWISG 674
R;Sinha, K.A.; Keen, J.K.; Ogun, S.A.; Holder, A.A.			Db	1196 --ADKSEKKEONHEKNEQIRIEDEVAKONKSN-----KAILD-----1229
Mol. Biochem. Parasitol. 76, 329-332, 1996			Qy	675 FEGKDKDGTGIVNLSKOTFIPVKFKEE--KKEEEENKPTFDVSKKDKDNPQV-----723
A;Title: Comparison of two members of a multigene family coding for high-molecular mass			Db	1230 -----TOLSEVEF-KICKFLKIDLRTMSDCLKETDIEKISNLSLTDQETKL 1278
A;Reference number: Z20507; PMID:97077455; PMID:8920022			Qy	724 -NHSQLNBSHKEKDQREBEHOKSSTDKDVATVLDKNNISSKTTNNPNK 773
A;Accession: T28676			Db	1279 ENKNIITLEKULESLKNOKNQKNDQKE-----LDEVNSKTKNIESNNVNO 1324
A;Status: preliminary; translated from GB/EMBL/DBBJ				
A;Molecule type: DNA				
A;Residues: 1-2401 <SIN>				
A;Cross-references: EMBL:U36927; PID:91041784; PID:91041785; PID: AAB41263.1				
R;Keen, J.; Holder, A.; Playfair, J.; Lockyer, M.; Lewis, A.				
Mol. Biochem. Parasitol. 42, 241-246, 1990				
A;Title: Identification of the gene for a Plasmodium yoelii rhoptry protein. Multiple co				
A;Reference number: A45521; MUID:91101660; PMID:2270106				
A;Accession: A45521				
A;Status: preliminary				
A;Molecule type: DNA				
A;Residues: 2260-2401 <KEE>				
A;Cross-references: GB: M34281				
RESULT 8				
T18440	hypothetical protein C0424w - malaria parasite (Plasmodium falciparum)			
C;Species: Plasmodium falciparum				
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999				
C;Accession: T18440				
R;Lawson, D.; Bowman, S.; Barrell, B.				
A;Status: preliminary; translated from GB/EMBL/DBBJ				
A;Molecule type: DNA				
A;Residues: 1-4550 <LAW>				
A;Cross-references: EMBL:Z98547; PID:e1325376; PID: e1325396; PID: CAB11121.1				
C;Genetics:				
A;Map position: 3				
A;Note: C0425w				
Query Match Score 230; DB 2; Length 2401;				
Best Local Similarity 21.0%; Pred. No. 0.031;				
Matches 187; Conservative 138; Mismatches 292; Indels 274; Gaps 46;				
Qy	2 LGBIAESKPKNLNGKEGSLLKDDTTGWB-HHHQNEESIKEKSSFTIDRNISTIRDFFNK 60			
Db	589 IDKITYDTDFENILTDKKTELETKFTGLSLNNHESNN--KEILTYFDKLANLGRKNEK-644			
Qy	61 DLCKLKKFREVDDFTSETGRMEEVYKYDDGNTIA-----YDDGTDLEYET 110			
Db	645 ---MLXQFNEKE-----KAVEDIKKONVDINKIVSNIETIYTISYNNINEDTNEI 693			
Qy	111 EK-----LDEBKXKTYGVLSPSKDHGFETLGK1SNSVSNQAKVYGN 151			
Db	694 GRSIELANTKYLEKVKANTNLNBEKEK-----KDYDFDQFGK---ERNIKYDPDEN 742			
Qy	5 TABSKPKNLNGKE--GSLKKDFTGV----EHHHOENEESIKEKSSFTIDRNISTIRD 56			

C; Genetics:
A; Note: PFC0345W

RESULT 10
A24594
probable
C;Species
C;Date: 2
C;Accessi
R;Holder,
Nature 31
A;Title:

A; Reference number: A24594; MUID: 86014355; PMID: 2995820

A; Accession: A24594

A; Molecule type: DNA

C; Superfamily: 1-1640 <HOL>

C; Keywords: major merozoite surface antigen

C; Surface antigen

Query Match Score 5.4%; Best Local Similarity 18.6%; Matches 167; Conservative 164; Mismatches 340; Indels 228; Gaps 39;

Qy 23 KDTG-VERHHQNEESIK-----EKSFTIDRNISTRDENKDLKLKKKFREVDD 75

Db 232 KDNVKGMDYIKKNGKTENINELIEEKKTIDKVNKNTKEEE-----KKLYQQY 283

Qy 76 FTSSTGKMEYDYKDGNTIAYDDSTDLEYETEKEDEIKSISKYIGWLSPSKDGHF 135

Db 284 DLSTYNGKOLEE-----AHNLISV-----LE-----KRDTLK-----KNNIKELL 319

Qy 136 GKLNSVSKNKAQTYGGNNYKSIEKATKVDHF-----SKTMFDLYANINDIVDGLAFA 188

Db 320 DKNBEIKNPPANSGNTNTLDRNKKEKEHEKEKEARTKXPNIDSFTDPL-----373

Qy 189 GDMRLFVUDNDQKKAIEKIRMPPEKIKETKSEY-----YVSSYGNV---IEJGE-GDJSK 239

Db 374 -ELEYYLREKN-KNIDISAKVETKESTEPEYNGTVPLSYNDINNALNELNSFGDLI- 430

Qy 240 NKPDLNLTWMEGSKYSDSEKQOYL-LKDNTIL-----RKGYALKVTYTNPGKTDML 289

Db 431 -NPFDYTREPSKNTYDNERKQTNEKEKIKI-BKKK-BSDRKSYEDRSKSLN----DIT 485

Qy 290 EG-----NGVSKB-----DIAKIQA-----NPNLRLSETTIVADSRNVNEDGRSTQSV 334

Db 486 KEYKEKLNEIYDSDKFNNNIDLTNFENKMGKRYSKYKVELTHPTFASEYENSKHNLKEKTK 545

Qy 335 LMSALDGF---NII-----RYCVTFKMDKGAEIDKGNLVTDSSKLVLFGDKDKEYT 385

Db 546 ALKYMEDSSLRNIVVEKELKYKKNLISIENETETLVE--NIKODEBQLF----EKKIT 598

Qy 386 GEKFENVAIKEKGDSMLIDTKEPVNLNSDKNYENPSENKSYKIVRNPFYLRSKISDGKF 445

Db 599 KDENKPKDEKILEYDIVKQVQKVLL--MNKIDELKKTQTLKRNVE--LKHNTHYVNSY 653

Qy 446 NWFLRVNNSVVDNYLIYGDHLHDNTNTRDENPIKL-----NWK-----480

Db 654 KOENKQE---PYLIVLKKEIDKLKVMPKYSVLINEEKNIKTQOQSNSBPESTEGL 709

Qy 481 -----DGDIMDGKMDWK-----ANGFPDKV 501

Db 710 TQGATTKGQGQAGSALESDVQAQAOBKRQAAQPVPVPEAKAQVPTPPAVNNKTEV 769

Qy 502 TDMDGNYLQGTGSDLNA-----KAVGHYOFYDNYKPEVNIDPKNTSTSLEYADGKSV 556

Db 770 SKD-----YLEKLYQFLNTSYICHYLYSHSTMNEKTLQYKITKEBSKUSSCDPLLL 826

Qy 557 FNIN-----DKRNGFDG---EIQEOHQHYINGKEYTSFNDIKQIIDDNLNIKIVV 603

Db 827 FNQNNITVYMSMFDLSLINSLSQLOMELYEKEMVCNLXKLKONDKTKLQNLBBAKYTSV 886

Qy 604 KOPARNTVKERFLINKDGTGEVSBLKPHRTVTIQLNGKEMSTTIVSEDFILPVYKBELEK 663

Db 887 KTLSSSSNQPLSLTPDKPEVSAANDTSHSTLNNSKLKENILS-----LGKKN 937

Qy 664 GTOFDGWFTISGPFGK-----KDACTGVINUSKDTFKKPVFKKIERKEEENKPTF-D 713

Db 938 IVO---ELIGQKSSSENYEKLKDSDFTYNFSFTNFTVRSKADDINSLNDESKRKLEED 993

Qy 714 VSKKKCDNFQVNHSOLNESHRKDLOREHSQSKSDSTKDVTATVLKNNISSK-SITNNP 771

Db 994 INLKCKTQLQSFOLNYKVKLKERLFLDKKTVGKYKMQIQUKUTLXEQLESKLNSLNNP 1052

SAZQK1

major merozoite surface antigen precursor - malaria parasite (Plasmodium falciparum)

(8).

C; species: Plasmodium falciparum

C; Date: 30-Sep-1987 #sequence_revision 30-Sep-1987 #text_change 09-Jun-2000

C; Accession: A25120

R; Mackay, M.; Goman, M.; Bone, N.; Hyde, J.E.; Scaife, J.; Cesta, U.; Stunnenberg, H.;

EMBO J. 4, 3823-3829, 1985

A; Title: Polymorphism of the precursor for the major surface antigens of Plasmodium fal-

A; Reference number: A91030; MUID: 3004972

A; Accession: A25120

A; Molecule type: DNA

A; Residues: 1-161 <MAC>

C; Comment: The merozoite stages of different strains have strain-specific surface antig-

C; Comment: P. falciparum has three stages: sporozoite, merozoite, and gameteocyte. The m-

C; Superfamily: major merozoite surface antigen

C; Keywords: glycoprotein; merozoite surface antigen; tandem repeat; transmembrane prot

F; 1-19:Domain:

F; 20-1631:Product: major merozoite surface antigen #status predicted <SG>

F; 67-84:Region: 3-residue repeats (S-G-T-P)

F; 1614-1631:Domain: membrane anchor #status predicted <MBN>

F; 97-259,755,759,835,911,955,1049,1156,1436,1563:Binding site: carbohydrate (Asn)

Query Match Score 5.4%; Best Local Similarity 18.1%; Matches 164;保守性 162; Mismatches 339; Indels 240; Gaps 37;

Query Match Score 5.4%; Best Local Similarity 18.1%; Pred. No. 0-062; Matches 164;保守性 162; Mismatches 339; Indels 240; Gaps 37;

Query 23 KDTG-VERHHQNEESTK-----EKSFTIDRNISTRDENKDLKLKKKFREVDD 75

Db 223 KDNVGMEDYIKKNGKTENINELIEEKKTIDKVNKNTKEEKKLQYAQYDLPFTYINKQ-----LTKKK 69

Qy 223 KDNVGMEDYIKKNGKTENINELIEEKKTIDKVNKNTKEEKKLQYAQYDLPFTYINKQ-----LTKKK 69

Db 70 FREVDDFTSETGRMEEYDYKDDKGNTIAYDDGTDLLEYETEKLDETKSKLYGVLSPSKID 129

Qy 283 LEPAHNLJSVLERKID-----LTKNNENIK-----307

Db 130 GHFEILGKISVSKNKAQVYYGGNNYKSIEIKATKVDHF-----SKTMTPDLYANINDIV 182

Db 308 ---ELDKINEIETKNNPANSGNTNTLDRNKKEKEKEIETKAKTIFKNDLSFTDPL 364

Qy 183 DGLAFAGDMRLFVKDNDQKKAIEKIRMPKEKIKETKSEY-----YVSSYGNV---IEGE 234

Db 365 -----ELEYYLREKN-KNIDISAKVETKESTEPEYNGTVPLSYNDINNALNEINS 416

Qy 235 -GDLSKNPKDPDNLTKMEGSKYTSDESKYQ-----LKDNTIL-----RKGYALKVTYNNP 283

Db 417 FGDLLI--NPFDYTREPSKNTYDNERKQTNEKEKIKI-EKKEKEIKERKFFINKEPSKNTYDNRKSKLN- 473

Qy 284 GKTDMLEG----NGVSK-----DIAKIQA-----NPNLRALSETTIVADSRVNEYD 328

Db 474 ---DITREKEYKLNLNEYIDSKFPNNNIDLTNFEROMGKRSYKEVLTHHTNFTASYENSKHN 530

Qy 329 RSTQSVLMSALDFG---NII-----RYQVFTPKMDGEALDQGNLVTDSSKVLFGK 379

Db 531 LEKLTALKYMDLSLRNIVVEKELKYKQNLSKINENETLVE--NIKIDDEQQL-----584

Db 380 DDKEYTGEDKVNVEAIEKGDSMLFDTKPVNLMSMDKNYFNPSKSNSKLYVVRNPBFYLRGKJ 439

Qy 585 -EKKITKDNEKDEKLEVSDFIVKQVQVQVLL--MNKIDELRKTQQLKQNV-----LKHNI 638

Db 440 SDKGGENWELRVNESVVDNYLIYGDHLHDNTRDNFL-----NVK-----480

Qy 639 HVPNSYQKENQ-----PYLIVLVLKGEIDKLKVMPKVESLNEEKKNIKTEGQSDNSEP 694

Qy 481 -----DGDIMDWGMKDYK-----AN 495

Db 695 STEGEITGQATTKPGQAGSALEGDSVQAQEQRQAQPPYPVVFPEAKAQVPTPPAPVN 754

Qy 496 GFPDKYTDMDGNYVLYOTGYSDLNA-----RAGVHYQFLYDNVKPEVNIDPKGNTSLEYA 550

Db 755 NKTENVSKLD---YLRLYEFNTSYTCHKYLVSHTSMNEKILQKVKITKEBEKSUSSC 811

Qy 551 DGKSKVYNIN-----DKRNGFDG---EIQEOHIIYINGKEYTSFNDIKQIIDKTL 597

Db 812 DPLDLJFNQNNIPVMYMSMFDLSLNSLSQLMETIYKEMVCNLYKLKNDKTKNULBEAK 871

Qy	598 NIKIVVKDFARNTTVKEFTILNKDTGEVELKPHRVTVITQNGKEMSTSIVSPEDFILPVY	657	Qy	335 LMSALDGF---NII-----RYQVIFTKMDKGAEIDKDGNLVTDPSSKLVLFGKDKEYT	385
Db	672 KVTSVKTLSSSSMQLSLTPQDKPEVSANDTSHSTNLNSLKLPEFTLS-----922		Db	546 ALKYMDESYSLNENIVVERELKYKNLISKIEETLVE--NIKDEEQLF---EKKIT	598
Qy	658 KGELEKGYQFDGWEISGPEGK-----KDAGYVINLISKDFTIKPVEKKEBENK 709		Qy	386 GEDKFVNTEAIKEDGSMLFIDTICKPVNLSMDKNYFPNSKSNSKLYVNRNPFYLRGKISDKGGF	445
Db	923 LGRNKNTIQ---ELIGGKSSSENFEYKILKSDTDFYNESFTNFVKSADDINSLNDESKR 978		Db	599 KDENKPDEKILEVSDFVKVQYQKVLL--MKNDIEELKKTOLIKNVE--LKHNIHVPNSY	653
Qy	710 PTF--DVSKKEDNPQVNHSQLEHESQKSDFSKDPTVATDKNNISSK-S 766		Qy	446 NWELRYNESVVDNYLITYGDLAIDNTDFNKL-----NVK	480
Db	979 KKLEEDINKLKTKLQLSLTDLYNKYKLUERLFDKKTGKVYKMQIKLTLKEQLESKLN 1038		Db	654 KOBNKQE---PYYLIVLKKEDLKVMPKVESLINEEKKNIKTEQQSDNSEPSTEGEI	709
Qy	767 TTNNP 771		Qy	481 -----DGDIMDWGMKDYK-----ANGFPDKV	
Db	1039 SLNNP 1043		Db	710 TQATTTKPGQOQGSALEGDSYQAQAOQKQOPPVPPVPAKAVQTPPPAPVNNTENV	769
Qy			Qy	502 TDMGNVYLQTYGSDINA----KAVGYHYQFLYDNVKPEVNDPKGNTSLEYADGKSV	556
Db			Db	59770 SKLD---YLEKLYEFLENTSYLICHKVLLVSHSTMNEKLLKQYKITCHEEESKLSSCPDLI	826
505603	major merozoite surface antigen precursor - malaria parasite - (Plasmodium falciparum) (strain 1.2)		Qy	557 FNIN-----DGRNNFGDQ--EIQEHIYINGKEYTSFNIDKLIQIDTLNIKIVV	603
N	Alternative names: gp195 surface antigen		Db	827 FNIQNNIPVMYSMFDISLANSLSQLNFSIYKEMVCNLYKLKDNDKLNLLBEAKVYSTV	886
C	Species: Plasmodium falciparum		Qy	604 KDFARNTTVKEPLANKDTGEVSELPKPHRTVTIONGKEMSTSIVSEDFILPVYKGELE	663
C	Accession: S05603 ; S04850		Db	887 KTLSSSNQPLSLTPQDKEPSANDDTSHSTNLNSLKLPEFTLS-----LGKKN	937
R	Myler, P.J.		Qy	664 GYQFDGWEISGPEGK-----KDAQGVINLISKDFTIKPVFKKIEKEEENKPTF-D	713
A	Submitted to the EMBL Data Library, April 1989		Db	938 IYQ---ELIGQKSENFYEXKLDSDPTFYNESFTNFVKSADDINSLNDESKRKLLEED	993
A	Reference number: S05603		Qy	714 VSKKDKNPQVNHSQLNFSHREHSDLQREHEHSKREDLQREHEHSKREDLQREHEHSKREDL	771
A	Accession: S05603		Db	994 INKLKTLQSLSPDLYNKYKLUERLFDKKTGKVYKMQIKLTLKEQLESKLNLSNNP	1052
A	Reference number: S04850 ; MUID:2663887				
A	Accession: S04850				
A	Molecule type: mRNA				
A	Residues: 1504-1639 <MYL2>				
A	Cross-references: EMBL:X15063				
A	Superfamily: major merozoite surface antigen				
C	Keywords: glycoprotein; merozoite; surface antigen				
P	1-1639 <MYL>				
F	1-19/Domain: signal sequence #status predicted <SIG>				
R	20-1639/Product: major merozoite surface antigen #status predicted <MAT>				
Nucleic Acids Res. 17, 5401, 1989					
A	Title: Nucleotide and deduced amino acid sequence of the gp195 (MSA-1) gene from Plas-				
A	Reference number: S04850 ; MUID:89345116 ; PMID:2663887				
A	Accession: S04850				
A	Molecule type: mRNA				
A	Residues: 1504-1639 <MYL2>				
A	Cross-references: EMBL:X15063				
A	Superfamily: major merozoite surface antigen				
C	Keywords: glycoprotein; merozoite; surface antigen				
P	1-19/Domain: signal sequence #status predicted <SIG>				
F	20-1639/Product: major merozoite surface antigen #status predicted <MAT>				
Query Match	5.4%		Score 218; DB 2; Length 1639;		
Best Local Similarity	18.6%		Pred. No. 0.062+; Mismatches 340; Indels 228; Gaps 39;		
Matches	167; Conservative 164; Mismatches 340;				
Qy	23 KDTTG-VENHHOENEESTIK-----EKSEPTIDENISTRDFENKLKLKIKKKFREVDD	75			
Db	232 KDNVKMEDIKNNKTTINNELLIESKKTIDKNKNATKEE-----KKKKLYQAYQ	283			
Qy	76 FTSETKRMEEYDYKDDKGNIAYDDGTDLYEYETEKULDEIKSKYGVLSPSKDGHPFIL	135			
Db	284 DLSIYNQKLEE-----AHNLISV-----LE---KRIDLK-----KNENIKEL	319			
Qy	136 GKISNVSKRAVKYVGNNYSIEIKATKDFH-----SKTMFDLYANINDIVGLAFA	188			
Db	320 DKNEIKNPPPANSNTPLLDKNNKKEEHEKEIKEAKTKEPNIDSFLTDPL-----	373			
Qy	189 GDMRLFVKONDOKKAIBIKTRMPEKIKETKSEYP---VYSSYGNV---IELGEG-DLISK	239			
Db	374 -ELEYLREKN-KNIDISAKVETESTPEYNGVTPSYNDSINNAINLNEINSFGDLI-	430			
Qy	240 NKPDLTCKMSEKGLYSDSEKQQL--LKDNIIL-----RKCYALKVTTYNPGKTM	289			
Db	431 -NPFDYTKEPSKNTYDNERRKKFINEIKERIKIERSKSYEDRSKSLN---DT	485			
Qy	290 EG----NGVYSKE---DIAKQKA---NPNLRASETTYADSRNVEDGRSTGSV	334			
Db	486 KEYKLINNETYDSEKNNNTDLTNFEMKGKRYSKYKVEKLTHNTFASTYNSKHNLEKLT	545			
Qy	335 LMSALDGF---NII-----RYQVIFTKMDKGAEIDKDGNLVTDPSSKLVLFGKDKEYT	385			
Db	546 ALKYMDESYSLNENIVVERELKYKNLISKIEETLVE--NIKDEEQLF---EKKIT	598			
Qy	386 GEDKFVNTEAIKEDGSMLFIDTICKPVNLSMDKNYFPNSKSNSKLYVNRNPFYLRGKISDKGGF	445			
Db	599 KDENKPDEKILEVSDFVKVQYQKVLL--MKNDIEELKKTOLIKNVE--LKHNIHVPNSY	653			
Qy	446 NWELRYNESVVDNYLITYGDLAIDNTDFNKL-----NVK	480			
Db	654 KOBNKQE---PYYLIVLKKEDLKVMPKVESLINEEKKNIKTEQQSDNSEPSTEGEI	709			
Qy	481 -----DGDIMDWGMKDYK-----ANGFPDKV				
Db	710 TQATTTKPGQOQGSALEGDSYQAQAOQKQOPPVPPVPAKAVQTPPPAPVNNTENV	769			
Qy	502 TDMGNVYLQTYGSDINA----KAVGYHYQFLYDNVKPEVNDPKGNTSLEYADGKSV	556			
Db	59770 SKLD---YLEKLYEFLENTSYLICHKVLLVSHSTMNEKLLKQYKITCHEEESKLSSCPDLI	826			
Qy	557 FNIN-----DGRNNFGDQ--EIQEHIYINGKEYTSFNIDKLIQIDTLNIKIVV	603			
Db	827 FNIQNNIPVMYSMFDISLANSLSQLNFSIYKEMVCNLYKLKDNDKLNLLBEAKVYSTV	886			
Qy	604 KDFARNTTVKEPLANKDTGEVSELPKPHRTVTIONGKEMSTSIVSEDFILPVYKGELE	663			
Db	887 KTLSSSNQPLSLTPQDKEPSANDDTSHSTNLNSLKLPEFTLS-----LGKKN	937			
Qy	664 GYQFDGWEISGPEGK-----KDAQGVINLISKDFTIKPVFKKIEKEEENKPTF-D	713			
Db	938 IYQ---ELIGQKSENFYEXKLDSDPTFYNESFTNFVKSADDINSLNDESKRKLLEED	993			
Qy	714 VSKKDKNPQVNHSQLNFSHREHSDLQREHEHSKREDLQREHEHSKREDLQREHEHSKREDL	771			
Db	994 INKLKTLQSLSPDLYNKYKLUERLFDKKTGKVYKMQIKLTLKEQLESKLNLSNNP	1052			
			RESULT 13		
			G90603		
			lipoprotein [imported] - Mycoplasma pulmonis (strain UAB CTIP)		
			C;Species: Mycoplasma pulmonis		
			C;Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 03-Aug-2001		
			C;Accession: G90603		
			R;Chambaud, I.; Heilig, R.; Perris, S.; Barbe, V.; Samson, D.; Galisson, F.; Moszer, I.		
			Nucleic Acids Res. 29, 2145-2151, 2001		
			A;Title: The complete genome sequence of the murine respiratory pathogen Mycoplasma pul-		
			A;Reference number: A95121; MUID:21267165; PMID:11353084		
			A;Accession: G90603		
			A;Status: preliminary		
			A;Molecule type: DNA		
			A;Residues: 1-1546 <KUR>		
			A;Cross references: GB:Al445566; PID:914090150; PID:GAC13908.1; GSPDB:GN00153		
			A;Experimental source: strain UAB CTIP		
			C;Genetics:		
			A;Gene: MYPU 7350		
			A;Genetic code: SG33		
			Query Match	5.4%	Score 217.5; DB 2; Length 1546;
			Best Local Similarity	20.2%	Pre#: No. 0.061;
			Matches	180; Conservative 125; Mismatches 336; Indels 249; Gaps 40;	
Qy	1 KLGIAESKPKNLNGKKGSLK-----KDTTGVEHHHQNEEESTKEKSSFTIDRNISTRD	56			
Db	103 KSKRLAEEQVGTSYNNEKFRLEAKNTKNTDQNSAKSQNIDSODHLKNSQKSNSN	162			
Qy	57 FENDLKLJIKKFRREVDDFTSETGRMMEYDVKYD-DKGNTIAYDGTDLYEYETEKLDE	115			
Db	163 IRNEQSKEQLTQNKEEHSANLQTQNLLESENTOQDLKPN-----NSNKEEKLKN	214			
Qy	116 IKSIKIYGVLSPSKDGHFPILGLK-TSNVSKAYVQYNNYKSTEIKATKXDFHSKTMFDL	174			
Db	215 VLS--FNKOEAVXGDSFERYFKEVSKLKEKVVDEITLDTKTFADF---SFEQ	267			

Qy	175 YANINDIVDGLAPAGDMRLFKNDOKKAETIKIRMPERKJETKSEPYVSSYGNVIEGLE	234	Db	266 LSEIKEYKDCKTTEISNSKRG--KDKIEFLEBKPKNEESNSNKVNINEINENIRNSEQY	322
Db	268 Y--IEKIKDLFNAST---IKDTYQ-----TNKIFLTNDEY-----298		Qy	52 -----	52 -----
Qy	235 GDLISKNKPDNLTKME----SGKIS-		Db	323 LKDIDAEBKQASTKVELPHKHBTTSNIKFKESEILGVETKSOKRINKAEDIMKEIERHNS	382
Db	299 --IKKNNNSNLKDKKKWPSNNFENEIYCYFLEXLDALVEIQALHKQVLEKNEIITGTGIVD	356	Qy	79 ETGKRMEEYDYKDDKGNI1AYDDGTDLEYETEKLDEIKSKYIGVLSPSKDGHEFILGKI	138
Qy	277 KVTTYNPGKTMDLEGNG-VYSKEDIAKI-----QKANPNURALS	314	Db	383 EJOTQVRGFOENLNKLNNEPHYDNAED-ELNNNDKSTNAKLVJETLIESVKIN--LSEI	437
Db	357 KYAAFIKSRELKSAKGILPSKDSOSAKITQTLINHILSRYPPEAPENULESLAKMGLINEL	416	Qy	139 SNSVSNAKVYGGNNYKSI-EIKATKVDHSKIM---TPDLYAN-INDIVDGLAPAGD	190
Qy	315 ETTIYADSRNVEDPGRST-QSVLMSALDGFNIIRYQVFTPKNDKGEBALDKGNLVTDSSK	373	Db	438 TNKQGEGEKIY-SKARDIMOKIKATSSENTAETLKVKDDQSNSVNVNQITT-----E	490
Db	417 KPEIYVEBGEVAYPTQDASINAQDQKIFD-----NKNLKDJK--SIVVDRN-	464	Qy	191 MRLFKDND-----QKKAETKIRMPERKJETKSEPYVSSYGNVIEGLEKPLBEGKLNKTSKINS	218
Qy	374 LVLFGKDDKEYGED--KFNVEAIKEDGSMLFDTKPVNLSMDKNGYFNPS-----421		Db	491 RNLIVTVEKNRNLGDTSTTINLEGALKESKGKVYSEKQOYLKDNLLRKGYALKV	550
Db	465 ITIPIAKSNVTITRKDSSSFTMPIVQRGALTFFIAPEPSOSINTNLGTSKDESSLVVKI	524	Qy	219 EYPVSSYYGNVIELEGDLSKNKPDLNTKOMESGKIVSDSEKQOYLKDNLLRKGYALKV	278
Qy	422 -KSNKLYYRNPEFLRGKJSDKGFNFNLRLRVNESVVNVLYIGDLH,DNTRDEN-----474		Db	551 TVGNFSSLFNNFDLNQDFKNKINDYENKM--GEYTNEFEGSLNKTSEN-----RNASENT	605
Db	525 EKQAKLVAKTGTAFINSSFSKYG-----SVFEN--YGSUVVIEGAKIWNVYSESG	572	Qy	279 TYNPGRKUDMLLEGNGYYSKEDI--AKIQKANPNRLALSETTIYASDRNVEDGRTSOSVL	335
Qy	475 -----IKLNVYDGDIMDWGMKDKYANGKANGFPDK-VTMGDNVYLQTCGDSLNAKAVG VH	525	Db	606 SDYNSAKTLLREA--QKEKVNLLNKEEANKYLR-----DVKKVESFR-----	646
Db	573 GITRNHYVQSSLTRNGE-----RDNTSTG--DKGIIYSGQNAISEGSDIDENKS-----620		Qy	336 MSALDGFIUIRQVFTKMDNGEADRGDNLUTDSSKLVLFGKDDKEYGEKDFKNEYAI	395
Qy	526 YQFLYDNVKPKEVNIDPKGNTSLEYADGKSVVFNINDKR--NNGPDGEIQBOHIYING-	580	Db	647 -----FIPNMK--ESLQKINEMIKEQSLTV-----NEGHGNVQLVENI	683
Db	621 FRSSLNL-E-KTNINFNSGSIVNNASVSKSILPIDSNSKIOISNNALNPFGSSAIP,KNN	679	Qy	396 KEDGSMMLFIDTKPVNLS-----MDKN-----YFNPDSKSNKLYIYVRNPEFYLRGKI	439
Qy	581 -----KEYTSNDIKOILDKLUNIKV-----602		Db	684 KB--LVDEN--NUSDILKQCATGKNEEIQKITHSTLKNKACTLGHVHDTSAKYVGSKI	736
Db	680 STMHLAGSLEKIKKEASORIEVTVLDFLTOAKLISPKNIIISLDNYQKLISSALFKFSVKN	739	Qy	440 SDKGGMFNPWLRVNESVVD-----NLYIYGDLHIDNTFDNPKLNVKD-----481	
Qy	603 VKDFFAR----NITVKEPTLNKDT--GEVSELKPHRTVTFIONGKEMSSSTIVSEE-DF	652	Db	737 TP----ELAITLELGDAKLUKTAQELKFESKNNVYLETENMSKNTNLEDVHKNIQDAYK	790
Db	740 INDFKHVLWNKTEKFPLWPDTKLFNTFYKTLQNDLIIQSGDFPSTEKTIKDLDLDF	799	Qy	482 -----GDMDWGMKDYKANGFPDKVTDMDGNVYI-----QFGYSDUNAKAVGH	525
Qy	653 -----ILPVYKGSELEKGYOF-DGWEISGFGEGK---		Db	791 VALELILHSDEIDTKQKD-----SSKLIEMGNQITLKVLINQYKONKISSLSKKEAVS	844
Db	800 YFRPTAAVKKLILTQLVRTIPKTHKEWAFAFEPNFDLQKWTETINELLRLDPFLFDIA	859	Qy	526 YQFLYDNVK-----PEVNIDPKGNTSLEYADGKSYVFNI-----NDKRNNFGDGELOEQ	574
Qy	690 DTFIKPVKIEKEKEENKPFDVSKKDNPOVNHSOLNEHSHREDIQR	739	Db	845 VRIC--GNVSKGHESLSKITCSKSYDNTIALEKQTELQNLRNSFTQETNTNSDSKLE--	900
Db	860 PERV-----ENGKMLPKPEY-IHTNWNVPVLEH-----FRNEDVAR	894	Qy	575 HIYNGKEYSNDIQLDITLNK-----VVKD-----605	
Db			Db	901 -----KIKTDFESLKNAL-KTLEGYNVALKASSDNHEHVQSKSEPVNPALSEKEETD	953
Qy			Qy	606 -PARNTIVKEFLNDKDTGEVSPSLKPHRTVTFIONGKEMSSSTIVSEE-DF	662
C;Species:	Plasmodium vivax		Db	954 IDSLNTALDELLKKGRCTCEVRYKLKIDKTVKEISDDTELNTIEKN-----VKAYLAYIK	1009
C;Date:	12-Mar-1993 #Sequence_revision 12-Mar-1993 #text_change 09-Sep-1997		Qy	663 KGYQFDGWEISGFGKDKDAGYVNLNSKDTFIPV-----FKKLKEKEEENKPTED-	713
R;Galinski, M.R.; Medina, C.C.; Ingravallo, P.; Barnwell, J.W.			Db	1010 KNYE-----DVTQD-----VLTNEHFNTKQVSNHEPTNPKNSSEETLTAVTDSK	1057
Cell 69, 123-126, 1992			Qy	714 --VSKKK-----DNPOVN-----HSOLNEHSHREDIQR	744
A;Title: A reticulocyte-binding protein complex of Plasmodium vivax merozoites.			Db	1058 TIIKLKGVIEVNENTMTESSAKEIEALYNELKONKTSLNELYTQTSNEVKLQEMKS	1117
A;Reference number: A42771; MUI:92315338; PMID:1617731			Qy	745 KSDSTDKVDTA--TVLD-----KNNISSKSTNN	770
A;Accession: B42771			Db	1118 NADKVIDVSKIFPNTVLDQKSNIVTNOHSINN	1149
A;Status: Preliminary: nucleic acid sequence not shown; translation not shown					
A;Molecule type: DNA					
A;Cross-references: GR:M88098; NID:gi160628; PID:gi160628					
A;Experimental source: Strain Belem, merozoites					
C;Genetics:					
A;Gene: RBP2					

RESULT 15

Query Match 5.4% Score 215.5; DB 2; Length 125;

Best Local Similarity 20.9% Pred. No. 0.056; Matches 207; Conservative 130; Mismatches 324; Indels 331; Gaps 50;

C;Species: Plasmodium falciparum C;Date: 22-Apr-1993 #sequence_revision 18-Nov-1994 #text_change 09-Jun-2000

Qy

C_Accession: A45597
 C_BiLi, W.B.; Bzik, D.J.; Tanaka, M.; Gu, H.M.; Fox, B.A.; Inselburg, J.
 C_Mol. Biochem. Parasitology: 46, 229-239, 1991
 C_Author: Characterization of the gene encoding the largest subunit of Plasmodium falciparum
 C_RefSeq: A45597
 C_Accession: A45597
 C_Status: Preliminary
 C_Molecule type: DNA
 C_Residues: 1-239 <L11>
 C_Note: sequence extracted from NCBI backbone (NCBIN:61099, NCBI:61101)
 C_Superfamily: Plasmodium DNA-directed RNA polymerase II large chain
 C_Keywords: nucleotidyltransferase; transcription

Query Match Score 214; DB 2; Length 239;
 Best Local Similarity 21.6%; Pred. No. 0.15; Gaps 50;
 Matches 191; Conservative 134; Mismatches 306; Indels 254; Gaps 50;

Qy	24	DTGVEHHQE-NEESKEKSSFTIDRNISTIRD--ENKDILKKKKPREVDDFTS 78
Db	1292	DNTYVEQLEMKELSKKNKTKKEQSKF----GTRDMHEDESEQMNKEETKKAK---FFI 1342
Qy	79	ETGKRMEEVDYKXDDKGNIAYDGTDLIEYETKLDETKSKY-----G 122
Db	1343	EK-KKGKRNHECNDDIEYNNTQD--NQYNNNSCNYKTSQNLENTHQNNDLSFTKNN 1398
Qy	123	VLSPSKDGH-----PEILGKISNSKNAKVYYGPKNSIEIKATPYDPHSKTMFDL 174
Db	1399	VILPPKEYHSIFAHFVNDRNVVLEKNUMLDKKFLANNEKNN--VVQSQSYNRMSKLNKKKI 1456
Qy	175	YANINDIVDG-----LAFAGDMRLFLVKNDQ--KKAELKIRMPBK---IETKSE 219
Db	1457	EI-INNIVRNEKEKCLRNRTKMDNDNNWSSDDDSIIAKKIKIKNNKEKRYHPKEEKEN 1515
Qy	220	YPVSSSYGNVIELEGDSLRNKEDNLTK-MESGKLYDSEKOYLKDNNIILRKGVALKV 278
Db	1516	PDR-NNYKMITDNNNHDV----NNLGV 1569
Qy	279	TTYNPGRKTDMLEEGNGPKT-----EGNGPKTDMLEEGNGPKT-----RALSETTIYADS-- 322
Db	1570	TNNN----NNYPNDNGYIKEETNNNELLTNNSMCNDNDFSEDEFNNINENDLITNKY 1626
Qy	323	-----RNVEDGRSTOSVILMSALDGENLIRYQVP--PTKMNDKGEALDKGNLVTDSKSL 374
Db	1627	YRQIFKNTIGFVSVFEYVESYKOHYILPPYEILKWTSPPLEYLETTIIFTNIFLHTKLSK- 1685
Qy	375	VLFCKD-----DKEYTGDEFKPVNKEAIKEDSMLFDTKPNLNSMDKNEYFNPSKSNKLYVRNP 431
Db	1686	----KEKPTPHQKNTGKMKIYIEETKK----WLFI---KAINI---YKYFFSKKSTEL---- 1728
Qy	432	EFYLRGKISDKGGFNWELRVNESVVDNLYIYGDLHIDNTRD-----KUNVKD 481
Db	1729	-----IKKDYFNYIIX-NYDISHRYIYH-DYSFINLKOLYLFIFIINYYKFV1STP 1779
Qy	482	GDIMDWGMKDYKANGFPDKVTDGMNVYLQTGYSIDLNAKAVGHYQFYLDNYKPEVNIDP 541
Db	1780	GD----AvesisaosigBGTQMTLTKTPHEAGVASHVY-TLGV-----PRIKEUINASN 1828
Qy	542	KGNTSIEYADGKSVFVNNDKRNFGF-----DGEIQHQHYINGKEYTS-----P 586
Db	1829	SICQPI----LNTPLVEONDNTNFALMMSKSKLEKTTIRDICMYIK-EDYTSRGVFLSVKF 1882
Qy	587	ND-----IKOLIDTLNI-----KI----VVKDOPARNTVK-EFLINKDT 621
Db	1883	NEELIQKLFLNINAYNIKDILKOSHINKIKINKIHINVKYKLHISLNDEDFIF---- 1938
Qy	622	GEVSELKPHRTVTIONGKEMSTSIVSSEDFILPVKGELERGYQFDGWEISGFECK-K 679
Db	1939	FQMESLKQSLDILYIYGDKDIRCLIKEEDI-----EVIDNEDEICD 1980
Qy	680	DAGYVINLNSKQDFIKPKVFKIEEK--BEENKEPTFDISKKDNQPVNHSQNEHRSKED 736
Db	1981	DMDEXXYNSQGT-----ELYERKCNSSKEENKNAIRKEEID-----DN 2019

Qy 737 LQRE-----HSQKS DSTKDVATVLDKNNLSSKSTNN 770
 Db 2020 LEKEENIIVYVSFKDSVNLKSEKKCIN----DDNNNNDDNNNN 2060

Search completed: January 28, 2004, 13:02:11
 Job time : 25 secs

GenCore version 5.1.6
 Copyright (c) 1993 - 2004 Compugen Ltd.
1 Protein - protein search, using SW model
 run on: January 28, 2004, 12:56:43 ; Search time 42 Seconds
 sequence: 1 KLGIEAESKPNLNGKEGS.....ATVLDKNNISSKSSTTNNNPNK 773
 scoring table: BLOSUM62
 Gapop 10.0 , Gapext 0.5
 searched: 830525 seqs, 258052604 residues
 total number of hits satisfying chosen parameters: 830525
 minimum DB seq length: 0
 maximum DB seq length: 2000000000
 post-processing: Minimum Match 0%
 Maximum Match 100%
 Listing first 45 summaries
 SPPMBP1 22.4

DR	PROSITE; PS50847; GRAM_POS_ANCHORING; 1.	DE Serine protease, subtilase family.
DR	PROSITE; PS00840; PA; 1.	GN SP0641.
DR	PROSITE; PS00137; SUBTILASE_HIS; 1.	GN Streptococcus pneumoniae.
DR	PROSITE; PS00138; SUBTILASE_SER; 1.	OS Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
DR	PROSITE; PS00678; WD_REPEATS; 1.	OC Streptococcus.
KW	Cell wall; Peptidoglycan-anchor; Protease.	OC
FT	NON_TER 1	NCBI_TaxID=1313;
SQ	SEQUENCE 2119 AA; 238226 MW; 5179B7F6B960A6A CRC64;	RN [1]
		RP SEQUENCE FROM N.A.
		PC STREPTOMYCE_TIGR4;
		RX MEDLINE=2157209; PubMed=11463916;
		RX TETEIN H.; Nelson K.E.; Paulsen I.T.; Read T.D., Tetelin H., Nelson K.E., Paulsen I.T., Read T.D.,
		RA Heidelberg J., Deboy R.T., Haft D.H., Dodson R.J., Peterson S., Kolonay J.P., Nelson W.C., Peterson J.D., Durkin A.S., Gwin M., White O., Salzberg S., Lewis M.R., Radune D., Umazza L.A., Holtzapfel E., Khouri H., Wolf A.M., Utterback T.R., Hansen C.L., McDonald L.A., Feldblyum T.V., Angiuoli S., Dickinson T., Hickey E.K., Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C., Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.; RT "Complete genome sequence of a virulent isolate of Streptococcus pneumoniae."
		RT Science 293:498-506(2001).
		RL EMBL; AB001373; AAC74791.1; -.
		DR MEROPS; S08_064; -.
Qy	1 KUGBIAESEKFKRNGLNGKRGSLKKDTGTYEHHHOENEEIKEKSFTIDRNISTIRDPEK 60	DR SP0641; -.
Db	1313 KUGBIAESEKFKRNGLNGKRGSLKKDTGTYEHHHOENEEIKEKSFTIDRNISTIRDPEK 1372	DR InterPro; IPR01899; Gram_pos_anchor.
Qy	61 DLKLKLKKFREYDDFTSETGRMEEEDYKDDKGNNIAYDGGTDLLEYETEKLEDEIKSKI 120	DR InterPro; IPR006192; LPXTG.
Db	1373 DLKLKLKKFREYDDFTSETGRMEEEDYKDDKGNNIAYDGGTDLLEYETEKLEDEIKSKI 1432	DR InterPro; IPR003137; PA.
Qy	121 YGVLSPSDGDFFELGKISNVSKNAKYVGNNSIEIKATYDFFSKMTMPDLYANIND 180	DR InterPro; IPR000209; Peptidase_S8.
Db	1433 YGVLSPSDGDFFELGKISNVSKNAKYVGNNSIEIKATYDFFSKMTMPDLYANIND 1492	DR InterPro; IPR001680; WD40.
Qy	181 IVDGLAPAGDMRLFVKNDQKRAEIKIAMPKEIKTKESEKPYVPSYGNVIELGEGLDSK 240	DR Pfam; PF00446; Gram_pos_anchor; 1.
Db	1493 IVDGLAPAGDMRLFVKNDQKRAEIKIAMPKEIKTKESEKPYVPSYGNVIELGEGLDSK 1552	DR Pfam; PF0225; PA; 1.
Qy	241 KPDNLTKMESCKLYSDSEKQQLKKDNILLRKGYALKUTTYNPKGKTDMLEGNGVYSKEDI 300	DR Prints; PF00052; Peptidase_S8; 2.
Db	1553 KPDNLTKMESCKLYSDSEKQQLKKDNILLRKGYALKUTTYNPKGKTDMLEGNGVYSKEDI 1612	DR Prints; PR00723; SUBTILISIN.
Qy	301 AKIQKANPNLRALESETTYIASRNVEDGRSTQSVLMSALDGENIIRYQVFTKNDKGEA 360	DR TIGRFANS; TIGR0167; LPXTG_anchor; 1.
Db	1613 AKIQKANPNLRALESETTYIASRNVEDGRSTQSVLMSALDGENIIRYQVFTKNDKGEA 1672	DR PROSITE; PS00847; GRAM_POS_ANCHORING; 1.
Qy	361 IDKDGNLVTDSSKVLFLGKDDKEYTEGDKFVNEAIKEGDSMLFIDTKPVNLSDKMYFNP 420	DR PROSITE; PS00840; PA; 1.
Db	1673 IDKDGNLVTDSSKVLFLGKDDKEYTEGDKFVNEAIKEGDSMLFIDTKPVNLSDKMYFNP 1732	DR PROSITE; PS00137; SUBTILASE_HIS; 1.
Qy	421 SKSMKIIYVRNPPEFYLRGKISDKGGFVNWLRLRVNESVVNDVLYIYCDLHINTTRDNIKLNK 480	DR PROSITE; PS00678; WD_REPEATS; 1; 1.
Db	1733 SKSMKIIYVRNPPEFYLRGKISDKGGFVNWLRLRVNESVVNDVLYIYCDLHINTTRDNIKLNK 1792	KW Protease; Complete_proteome.
Qy	481 DGIDMDWGMKDYZANGFDPKVTDGMNTQTYSDLNAAKAVGHYQPLYDNVKPEVNTID 540	SEQUENCE 2140 AA; 240426 MW; FA44AD8E2938B334 CRC64;
Db	1793 DGIDMDWGMKDYZANGFDPKVTDGMNTQTYSDLNAAKAVGHYQPLYDNVKPEVNTID 1852	Query Match 100.0%; Score 4026; DB 16; Length 2140;
Qy	541 PKGNTSIVVEADGKSVVPNINDKERNNGFGEIQEHOIHYINGKEYTSFNDIKQIDKTLNIK 600	Best Local Similarity 100.0%; Pred. No. 6.6e-154; Mismatches 0; Indels 0; Gaps 0;
Db	1853 PKGNTSIVVEADGKSVVPNINDKERNNGFGEIQEHOIHYINGKEYTSFNDIKQIDKTLNIK 1912	Matches 773; Conservative 2140 AA; 240426 MW;
Qy	601 IYVYDFARTNTVKEFILNKDTGEVSELKPHRVTVTIONKGEMSSIVSEDFILPVYKGE 660	Qy 1 KLGELIAESKFKRNGLNGKEGSLKKDTGTYEHHHOENEEIKEKSFTIDRNISTIRDPEK 60
Db	1913 IYVYDFARTNTVKEFILNKDTGEVSELKPHRVTVTIONKGEMSSIVSEDFILPVYKGE 1972	Db 1334 KLGELIAESKFKRNGLNGKEGSLKKDTGTYEHHHOENEEIKEKSFTIDRNISTIRDPEK 1393
Qy	661 LEKGYQFDWEIFSEGKDDAGYVINSKDTFKPKVFKKIEEKKEEENKPTDVSKKDN 720	Qy 61 DLKLKLKKFREYDDFTSETGRMEEEDYKDDKGNNIAYDGGTDLLEYETEKLEDEIKSKI 1453
Db	1973 LEKGYQFDWEIFSEGKDDAGYVINSKDTFKPKVFKKIEEKKEEENKPTDVSKKDN 2032	Db 1394 DLKLKLKKFREYDDFTSETGRMEEEDYKDDKGNNIAYDGGTDLLEYETEKLEDEIKSKI 1453
Qy	721 PQVNHSQLNEBRSKEDLOREEHRSOKSDSTKDTVATVLKNNISSKSTNNPNK 773	Qy 121 YGVLSPSKGDGFEPILGKISNVSKNAKYVGNNSIEIKATYDFFSKMTMPDLYANIND 180
Db	2033 PQVNHSQLNEBRSKEDLOREEHRSOKSDSTKDTVATVLKNNISSKSTNNPNK 2085	Db 1454 YGVLSPSKGDGFEPILGKISNVSKNAKYVGNNSIEIKATYDFFSKMTMPDLYANIND 1513
Qy	301 AKIQKANPNLRALESETTYASRNVEDGRSTQVLMALDGFVNIRYQVFTPMDNGKSEA 360	Db 1574 KPDLNLTKMEGKLYSDSEKQYLKDNILRKGYALKVTTYNFGKTDMLEGNGVYSKEDI 1633
Db	1634 AKIQKANPNLRALESETTYASRNVEDGRSTQVLMALDGFVNIRYQVFTPMDNGKSEA 1693	Qy 361 IDKGNLIVTDDSSKVLFLGKDDKEYTGDKFVNKAIEKGDSMLFIDTKPVNLSDMKNYFNP 420
Qy	361 IDKGNLIVTDDSSKVLFLGKDDKEYTGDKFVNKAIEKGDSMLFIDTKPVNLSDMKNYFNP 420	Db 1694 IDKGNLIVTDDSSKVLFLGKDDKEYTGDKFVNKAIEKGDSMLFIDTKPVNLSDMKNYFNP 1753
Db	1694 IDKGNLIVTDDSSKVLFLGKDDKEYTGDKFVNKAIEKGDSMLFIDTKPVNLSDMKNYFNP 1753	RESULTS 2
Q9TRY6	PRELIMINARY; PRT; 2140 AA.	Q9TRY6 PRELIMINARY; PRT; 2140 AA.
AC		AC Q9TRY6;
DT	(TREMBLrel. 18, Created)	DT 01-OCT-2001 (TREMBLrel. 18, Last sequence update)
DT	01-MAR-2003 (TREMBLrel. 23, Last annotation update)	DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)

Best Local Similarity 94.7%; Pred. No. 1.5e-144;		Matches 732; Conservative		Indels 1; Gaps 1;	
Qy	Ay	Qy	Ay	Qy	Ay
421	SRSNKLYVNPFFYLRSKISKDKGGFWNLRLNYESVVDNYLTYGDLHIDNTDRDNIKLNVK	480			
1754	SRSNKLYVNPFFYLRSKISKDKGGFWNLRLNYESVVDNYLTYGDLHIDNTDRDNIKLNVK	1813			
481	DGDIMDGWGMKDYKANGEPDKYTDMDGVYQIOTGYSIDLNAKAVGHYOFLYDVKPENVID	540			
1814	DGDIMDGWGMKDYKANGEPDKYTDMDGVYQIOTGYSIDLNAKAVGHYOFLYDVKPENVID	1873			
541	RKGNTSIEYADCKSYVNPNIKRNNGFGEIQEQQHYINGKEYTSNDIKQIDIKLNK	600			
1874	RKGNTSIEYADCKSYVNPNIKRNNGFGEIQEQQHYINGKEYTSNDIKQIDIKLNK	1933			
601	IIVVKDFARNTTVKEFLINKDTGSESELKPHRTVTIONKGNSSTIVSEEDFILPVYKGE	660			
1934	IIVVKDFARNTTVKEFLINKDTGSESELKPHRTVTIONKGNSSTIVSEEDFILPVYKGE	1993			
661	LEKGYQFDGMWLSGFEGKKDAGYVITNLSKDTPIKPVFKKEEEFNKPTEDVSKKKDN	720			
1994	LEKGYQFDGMWLSGFEGKKDAGYVITNLSKDTPIKPVFKKEEEFNKPTEDVSKKKDN	2053			
721	FQVNHSQNLNESHRKEDLQREHSQKSSTDSTKVATVLKNNISSSKSTTNNPK	773			
2054	FQVNHSQNLNESHRKEDLQREHSQKSSTDSTKVATVLKNNISSSKSTTNNPK	2106			
RESULT 3					
Q9S4M8	PRELIMINARY;	PRT;	2144 AA.		
ID	Q9S4M8				
AC	Q9S4M8;				
DT	01-MAY-2000 (TREMBLrel. 13, Created)				
DT	01-MAY-2000 (TREMBLrel. 13, Last sequence update)				
DT	01-MAR-2003 (TREMBLrel. 23, Last annotation update)				
DB	Cell wall-associated serine proteinase PrtA precursor.				
GN	Streptococcus pneumoniae				
OS	Bacteria; Firmicutes; Lactobacillales; Streptococaceae;				
OC	Streptococcus.				
NCBI_TaxID	1313;				
OX					
RN	SEQUENCE FROM N.A.				
RC	STRAIN-3_R;				
RA	Bethé G., ten Thoren E., Bongarts R.J.M., Heinz H.-P., Zysk G.;				
RT	"Cloning and sequencing of a novel surface protease of Streptococcus pneumoniae."				
RL	Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.				
-!	SUBCELLULAR LOCATION: ATTACHED TO THE CELL WALL PEPTIDOGLYCAN BY AN AMIDE BOND (BY SIMILARITY).				
EMBL	AF127143; AAD48399.1; -.				
HSSP	P00782; 2SBT.				
DR	MEROPS; SCB 064; -.				
DR	InterPro; IPR001899; Gram_pos_anchor.				
DR	InterPro; IPR006192; LPXTG.				
DR	InterPro; IPR003137; PA.				
DR	InterPro; IPR00209; Peptidase_S8.				
DR	Pfam; PF00082; Peptidase_S8; 2.				
DR	PRINTS; PR00723; Subtilisin.				
DR	TIGRFAMS; TIGR0167; LPXTG anchor; 1.				
DR	PROSITE; PS00847; GRAM_POS_ANCHORING; 1.				
DR	PROSITE; PS00840; PA, 1.				
DR	PROSITE; PS00137; SUBTILASE_HIS; 1.				
DR	PROSITE; PS00138; SUBTILASE_SER; 1.				
DR	PROSITE; PS00618; WD_REPEATS; 1.				
KW	Cell wall; Peptidoglycan-anchor; Signal.				
FT	SIGNAL_19				
FT	CHAIN	20	2144	PRTA.	
SEQUENCE	2144 AA;	240724 MN;	Score 3793.5;	DB 2;	Length 2144;
SQ					

DeJoff B.S.,	Bistrem S.T.,	Fritz L.,	Fu D.J.,	Fuller W.,	Geringer C.,
Gilmour D.J.,	Lee L.N.,	Khatri H.,	Krafft A.R.,	Lagae R.E.,	
LeBlanc D.J.,	McNamee M.,	Elefante F.J.,	Ilu J.,	Matsuhashi P.,	
McAhern S.M.,	McNamee M.,	McLawsster K.,	Mundy C.W.,	Nicas T.I.,	
Norris F.H.,	O'Gara M.,	Perry R.B.,	Robertson G.T.,	Roche P.,	
Sun P.-M.,	Winkler M.E.,	Yang Y.,	Young-Bellido M.,	Zhao G.,	
Zook C.A.,	Baltz R.H.,	Jaskunas S.R.,	Rosbeck P.R. Jr.,	Skatrud P.L.,	
Glass J.I.,	"Genome of the bacterium Streptococcus pneumoniae strain R6."				
RA	RA	RA	RA	RA	RA
DR	DR	DR	DR	DR	DR
SQ	SQ	SQ	SQ	SQ	SQ
Query Match	92.1%	Score 3709.5;	DB 16;	Length 2144;	
Best Local Similarity	92.4%	Pred. No. 3..e-11;			
Matches 714;	Conservative	26; Mismatches 32;	Indels 1;	Gaps	
Qy	1	KLGTAESPKFLNGLNGKEGSLLKDTGVEHHHOBNEETIKEKSFTIDRNNTSTIREDENK	6		
Ddb	1339	KLGTESEDKLKNAKSPEENT -NNNQARDEDSPDEKSVEVGASLEINKTSTIREPENK	1		
Qy	61	DLKLKLKKKFREYDDFTSETGKMEBEYDKYDDKGNTIAYDGTDLLEYTERKLDEIKSKI	1		
Ddb	1398	DLKLKLKKKFREYDDFTSETGKRIEYDKYDDKGNTIAYDGSALOYETEFDEIKSKI	1		
Qy	121	YGVLSPSKDGFELIGKISNVSKNAKTYGGNNKSIETKATYDFNSKTMPTDLYANIND	1		
Ddb	1458	YGVLSPSKDGFELIGKISNVSKNAKTYGGNNKSIETKATYDFNSKTMPTDLYANIND	1		
Qy	181	IVDGGLAFIGDMRLFVKDQKAEITKTRMPKETKETSEYPVSSYGVIEGEGDLSKN	2		
Ddb	1518	IVDGGLAFIGDMRLFVKDQKAEITKTRMPKETKETSEYPVSSYGVIEGEGDLSKN	1		
Qy	241	KPDNLTKMMSGKYSDESKQQYLKDNTILRKSGYALKYTYYNGKTDMLGEKGVSKEDI	3		
Ddb	1578	KPDNLTKMMSGKYSDESKQQYLKDNTILRKSGYALKYTYYNGKTDMLGEKGVSKEDI	1		
Qy	301	AKIQKANPILRALSSETTIADSRRNVEDRSTSQVYLMALDGMNIRYQVFTKMDNGEAA	3		
Ddb	1638	AKIQKANPILRALSSETTIADSRRNVEDRSTSQVYLMALDGMNIRYQVFTKMDNGEAA	1		
Qy	361	IDKDGNLVTDSSKLVLFLGKDDKEYGEDKFNVAIKEKGDSMFLIDTCKPVNLSDMDKYNFNP	4		
Ddb	1698	IDKDGNLVTDSSKLVLFLGKDDKEYGEDKSNVRAIKEKGDSMFLIDTCKPVNLSDMDKYNFNP	1		
Qy	421	SKSNKIVYRNPNPEFYLRGKTSIDKGFFNWELRVNESVVNDYLITYGDLHIDNTRDENTKLNUKV	4		
Ddb	1758	SKSNKIVYRNPNPEFYLRGKTSIDKGFFNWELRVNESVVNDYLITYGDLHIDNTRDENTKLNUKV	1		
Qy	481	DGDIMDWGMKDYKANGFPDKYTMGNTVINDKRNGFDGEIQCQHYQFLYDNVKPEYNID	5		
Ddb	1818	DGDIMDWGMKDYKANGFPDKYTMGNTVINDKRNGFDGEIQCQHYQFLYDNVKPEYNID	1		
Qy	541	PGKNTSIEYADGKSVVNINDKRNGFDGEIQCQHYQFLYDNVKPEYNID	6		
Ddb	1878	PGKNTSIEYADGKSVVNINDKRNGFDGEIQCQHYQFLYDNVKPEYNID	1		
Qy	601	IVVKDFARNTTVKEFILNDTGGEVSELKPHRYVTTIONGKEMSTIVSEEDFLPYVKGE	6		
Ddb	1938	IVVKDFARNTTVKEFILNDTGGEVSELKPHRYVTTIONGKEMSTIVSEEDFLPYVKGE	1		
Qy	661	LEKGYQFDGMWISGPEGKDAGYINLNSKDTFKPVFKKIEEKKEENPKTDVSKKDDN	7		
Ddb	1998	LEKGYQFDGMWISGPEGKDAGYINLNSKDTFKPVFKKIEEKKEENPKTDVSKKDDN	2		
Qy	721	PQVNHSQLNESHRKDLOREEHRSOKSDSTKDVTAVLDKNNISSKSTTNNPBK	773		
Ddb	2058	PQVNHSQLNESHRKDLOREHSOKSDSTKDVTAVLDKNNISSKSTTNNPBK	2110		

RESULT 5
Q8TB84

Oliver K., Ormond D., Price C., Quail M.A., Rabinowitzsch B., Rajandream M.A., Rutter S., Rutherford K.M., Sanders M., Simmonds M., Sharp S., Smith R., Squares R., Stevens K., Seeger K., Taylor A., Tivey A., Unwin L., Whitehead S., Woodward J., Sulston J.E., Craig A., Newbold C., Barrell B.G.; "Sequence of Plasmodium falciparum chromosomes 1, 3-9 and 13."; Nature 419:527-531(2002). EMBL; AL929353; CD51583.1; -.	Db	1720	-ISYKMLKNN -QINDTYKLKNRSLSTVELJKQ -DIKFLNEDVLKKKEMVTLS 1777
Best Local Similarity 20.7%; Pred. No. 0.019; Matches 186; Conservative 152; Mismatches 324; Indels 44; SEQUENCE 1777 AA; 213320 MW; 244467CF190C522 CRC64;	Qy	4	EIAESKFKNLNGNGKEGLKDDTGYEHHOE-----ESTIKERSSFTIDRNINISTRDF 57 977 ELDDDKKKL--DEENELLDKKR-KRLDENELLDDKKRKL-BE 1030
Query Match Score 255; DB 5; Length 177;	Dy	58 ENK---DLKKL-----KCKFREYDDFTSETGKRMEEYDKYDKDGNITAYDDGT 104 1031 ENELDDDKKKLDEENELLDDKKRKL-BE 1090	
Best Local Similarity 20.7%; Pred. No. 0.019; Matches 186; Conservative 152; Mismatches 324; Indels 44; SEQUENCE 1777 AA; 213320 MW; 244467CF190C522 CRC64;	Db	105 DLEYETDEIJKS---GVGLPS-KDHF-----ELKGKSNVSKRAK-----146 1091 PIOPHEHNLNEKTKTQIODEYNITIAELEMHSKSKILLETKIKIENDENDIKRKOSQIEN 1150	
Query Match Score 255; DB 5; Length 177;	Dy	147 VYGGNNYKSTEIRATKATKDYFHKSKMTDFDLXANINDIVDGLAFAGDMRLFVDNDQKAIEK 206 1151 TYRNSMMDNINSYKSUVKTKTFNFNDNIEE-----1184	
Best Local Similarity 20.7%; Pred. No. 0.019; Matches 186; Conservative 152; Mismatches 324; Indels 44; SEQUENCE 1777 AA; 213320 MW; 244467CF190C522 CRC64;	Db	207 IRMPBEKIKETKSEPYVSSYGNVIELGDSLKNPKDNLTKMESGKITYSEKQQYLLKD 266 1185 ---EDKITEQKENTKLYTDDISNM-----LTKNKKSSLYSINSEKDLTNEQELYSNN 1234	
Query Match Score 255; DB 5; Length 177;	Dy	267 NI-----ILRGALVKVITYNPGRDMLGGVTSK----EDIAKIQK-ANPN 309 1235 NTENNDIEKDLTNMEKKNNNNNISYKKCOSNIYDINSNNKPLMDTESLIKIQNDINEK 1294	
Best Local Similarity 20.7%; Pred. No. 0.019; Matches 186; Conservative 152; Mismatches 324; Indels 44; SEQUENCE 1777 AA; 213320 MW; 244467CF190C522 CRC64;	Db	310 LRALS---ETTIYADSRRNVEDGRSTQSVL-SALDGENTIITYQVFHKMDKGAEADKDG 366 1295 VKLQKREIYEKEKFNFKLNLKENVNKNKIKDR--ENELNKEEELKRMKE 1352	
Query Match Score 255; DB 5; Length 177;	Dy	367 LVTDSKLVLF-----GKDKKEY-----TGDKFNTVAIKEGDSMLFDTKVNLSM 413 1353 ILLSKENIINYSLNSSKINKNKLDEYTKGKBDINSTMKSFKSHL-SDTN -NIQP 1409	
Best Local Similarity 20.7%; Pred. No. 0.019; Matches 186; Conservative 152; Mismatches 324; Indels 44; SEQUENCE 1777 AA; 213320 MW; 244467CF190C522 CRC64;	Db	414 DKNFNPNSKNIYTRNPFFLRLGKISDKGGFENWEIRV-----NESTVDNVLILY-G 463 1410 FKNDIN--NDKLYEDNISYY-GNNLSSHDNMHNKMYDNLSCDNNMSHDNNMICHSN 1465	
Query Match Score 255; DB 5; Length 177;	Dy	464 DLHIDNTDFNKLUNV-----KDGIDMDWGMKDYKANGFPDKUTMDGNYV-----509 1466 NISTDNNPNTNTHMKLSELBNNEFNKNKNDY---YDIDNFDRNNTIHNNNNISKI 1521	
Best Local Similarity 20.7%; Pred. No. 0.019; Matches 186; Conservative 152; Mismatches 324; Indels 44; SEQUENCE 1777 AA; 213320 MW; 244467CF190C522 CRC64;	Db	510 ---LOTGSDLNAKAVGTVHQFLYDNVKPEVNIDPGKNTSTEYAGKSVYFNINDKRNG 566 1522 SNDSTSGNMNNE-----LHDNL-----NSLSEKEIMYKSLIKSRDSEINT 1562	
Query Match Score 255; DB 5; Length 177;	Dy	567 FDGE QEO-----HIVYNGK-----EYTSFNDKQI1DRTLN 598 1563 LKLTEKQRTKEEKWNQNINQNL DISSDMNINYINGNISNGVNYDTSMHYLELPERKFLN 1622	
Best Local Similarity 20.7%; Pred. No. 0.019; Matches 186; Conservative 152; Mismatches 324; Indels 44; SEQUENCE 1777 AA; 213320 MW; 244467CF190C522 CRC64;	Db	599 IKIVYKDFEARNNTTKEFLNKDGT---EVYSLKLPHRVVTIIONGKEMSTIUSSEDFI -- 653 1623 LS-EYED--NEKMSILSNELNLSVLYKEINNIK-----EYNINVLRKNEFIGN 1667	
Query Match Score 255; DB 5; Length 177;	Dy	654 LPVYKGELEKGYQFDGWLSGFEGKDKDAGYVINLSKDTFIKPVFKKEKEENKPTF 712 1668 LLLNLFNDLNNYRL--KENYKEKEVHSQIIISERDFKELQNTNLBCKLKE --- 1719	
Best Local Similarity 20.7%; Pred. No. 0.019; Matches 186; Conservative 152; Mismatches 324; Indels 44; SEQUENCE 1777 AA; 213320 MW; 244467CF190C522 CRC64;	Db	374 LYLFKGDKKEYGDKFENVEA1KED--GSMFLPDTKPVNLSMKDNFV--PSKSNKLY 424 593 ---DKLDEYFSNEKTDYVTFNKGFLGNLDSINKLNNNDQKEYINSKIDSKRNELS 644	
Query Match Score 255; DB 5; Length 177;	Dy	428 VRNPETY-----LKGKISDKGGFENWELRNESVDNYLIYGLDHLIDNTRD 477 649 TMEDDIFNAKQFASITNNNENISNKIKD-----LNEF1SNEDSSKEL-LDEIRK 690	
Best Local Similarity 20.7%; Pred. No. 0.019; Matches 186; Conservative 152; Mismatches 324; Indels 44; SEQUENCE 1777 AA; 213320 MW; 244467CF190C522 CRC64;	Db	473 FNKLNVKDGDIMDWGMKDYKANGFPIKVTDMDGNNV-YLQTGYDLDNAKA-----V 524	

Db	699 YKQQFD- KIKDAVANTEVSPE- NTLOKIDSIKSINELNTNAYDIINTKLDDKLUNY 756	Db	859 -----NESPLKMMKONY-----TEHYDDKKNNIPLVKNEYDTYSNIHNGI 900
Qy	523 GIVYQFLDNVKP-----EVNTDEKGNTSIEYAD-GRKVVENINDKRNGFDEGEIQEIGHI 576	Qy	180 -----DIVDGLAFAGDMR----LFWVKDN----DOKK----- 202
Db	757 GSEFKLNLYNASDLIDDTORNRNDERKVQNLNEYLKRNQSIEINDIVNN----- 805	Db	901 HENSMILKKNYKCACTPHGYSRNHQKRNQYSHHRNGTVLKPLVNTNN 960
Qy	577 YINGKEYTSFNDIKOIIDTKLNKIVVWDFAARTTVKEFILNKDGTGEYSE-- LKPHRVT 633	Qy	203 -----AEKIRMPKIKETKS --EPYVSSYGVIELEGDSKPKDNLTKMESGKI 253
Db	806 FI- KELKTFNNT- -TNSLNL- ELLTD- -DINDKIFKLYRELNKISTNNLKLKY- 855	Db	961 VAVNEFADINLSAQKRLHSLSKMGYEDSKMENTRNKLYNNINNNNNNNNDN -----NI 1013
Qy	634 VTIONGKEMSTTIVSEDFILPVYKGEBGYQDFGWELSGFGRKDGAYVINVNLSDKF- 692	Qy	254 YSDSEKQY-- LLKDNTLIRKGAL----KVTYTNGKTDMLEGNVSYSEDIAK--- 302
Db	856 NEIDNVNEKLSTIVENLQFINSFUSIENPQG-----SITSH-----INFLNTLA 900	Db	1014 YNDNEYCOYNNSYCFDISDLKMPFLHQNSKLTHSNKNSFFNGINVESXHHLANPEI 1073
Qy	693 -IREVFKEKICBKEBEENKPTEDYSKCKKONPQVNHSRAXEDLOREBEHSOKSDSTKD 751	Qy	303 -----IQRKAP----- 309
Db	901 GINDVLNKLNLKIMADTTRGDTNIR-----DEIRKQISSENIKSQKFNEKNE -KD 950	Db	1074 KTPAHNSYPIILNOGLINNPLOCGLYDSNQRKNHNVVYIKRNBNLNKNGISTNLVLRKG 1133
Qy	752 VTATV-----LDKNNISSKSTTN 770	Qy	310 LRALS-----ETIYADSRSNVEDGRSTQSVLMSALDGFNTIRYQVTFKMD 356
Db	951 LKRLISFFNDKLKNYCNIASAGYTEYN 974	Db	1134 LRKXISTINGKPFESFSNMDNKNONTMEGLINQDNVNWNKNEC- DNKHMRTRTSLSNFVSR 1192
Qy	DBI_081121_ PRELIMINARY; PRT; 2227 AA.	Qy	357 K-GE--AIDDGNLVTDSSKLVLFGK-DDKEVYGDKENVEAIKD-GSMLFLD---TKP 408
AC	DBI_081121;	Db	1193 SYGEHKSLDVYQECYVKNKL1 -NKONDKCY -EDNNNSYLNEDDNASMQYEETNSNP 1248
DT	01-MAR-2003 (TREMBLrel. 23, Created)	Qy	409 VNLSMD--KTFGNPSKSNNKNTVNPFFYLRLKGKISDKGFWME----LRYNEV---- 455
DT	01-MAR-2003 (TREMBLrel. 23, Last sequence update)	Db	1249 YIVDQEENRMKVN-----NVLTNNNSNYVDSKRNDSKSENKSDDILNNENIHTLKD 1304
DE	Hypothetical protein.	Qy	456 -----VDNLYLIGDLHIDNTRDPDENIKLNVKGDDIMW--GMKDQXANGFPDKV-TDMDG 506
GN	PF11_0554.	Db	1305 QKKKIONNNEFISEOADETENINSQEEVYKEHEPL-WVINASNEEKSYEEELIYSDMSS 1363
OS	Plasmodium falciparum (isolate 3D7).	Qy	507 NVLQTGYSIDLNAKAVGHYQLDNYV/KPEVNIDPKGNTSLEYADGKSVVFNINDKANG 566
OC	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.	Db	1364 NRVTKVKYSDMRNNEVNLINEDNLITTEKYKVOLEKENKMDIMYETVEENINTKTENTND 1423
NCBI_TaxID	=36329; [1]	Qy	567 FDFEIQQHIIYNGKEYTSFNDIKOIIDTKLNKIVVTDKFDANT--TVK-BFLNKKDTG 622
RN	RESTRNACE FROM N_A.	Db	1424 INEEVR-----NEQRKBSNHNID---TNINHIDEYPNDTINFNIKIDECVHNENN 1472
RC	STRAIN=3D7;	Qy	623 EVSELKPCHRVTYTQNGKEMSTSIVSEDFDILPVYKGELKEYQFDGWBISGFEKGKDAQ 682
RX	MEDLINE=22255700; PubMed=12368864;	Db	1473 MYNSIEQTYFVHDTRNNHLVDK--NRQNFI-----FEEGELNELNFEKK-V 1516
RA	Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W., Carlton J.M., Pain A., Nelson K.B., Bowman S., Paulsen I.T., James K., Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S., Chan M.-S., Nene V., Shallow S.J., Sub B., Peterson J., Angiuoli S., Pertea M., Allen J., Selengut J., Haft D., Mather M.W., Vaidya A.B., Martin D.M.A., Fraunfelder M.J., Fraunfelder M.J., Roos D.S., Ralph S.A., McFadden G.I., Cummings L.M., Subramanian G.M., Mungall C., Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W., Fraser C.M., Barrell B.; RTR "Genome sequence of the human malaria parasite Plasmodium falciparum"; RFL EMBL: AE014941; LAN35338.1; -.	Qy	683 YVNLNSKD-----TFLKPVFKKIEBKKEENKPTF-----DVSRRKDDN 720
RK	Hypothetical protein.	Db	1517 YIENNTRDHKGDSTSNTSLRNICKSENDNEKENNTYVVKGEKGKIKRKVSMKCRN 1576
SEQUENCE	2227 AA;	Qy	721 PQVN-HFQLN-----BSHREKDQREBHSOKS 746
RQ	262840 MW;	Db	1577 EKNEENYINNIYDMDNHQNDTICKENDDEEN 1609
RESULT 10			
Qy	6.3% Score 252.5; DB 5; Length 2227; Best Local Similarity 19.1%; Pred. No. 0.031; Matches 190; Conservative 155; Mismatches 319; Indels 329; Gaps 49;	Q8IHY4	PRELIMINARY; PRT; 2849 AA.
Qy	10 FKNLNGK-----GSJLKDDTGVEH-----HHQZNEEESIKEK 42	ID Q8IHY4; AC Q8IHY4;	CREATED
Db	690 YANNNNNKGTCISNIKISENKYPFKNSYMTNNTLPHRNATTSNRRNEEVEKEK 749	DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)	DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
Qy	43 SSFTIDRNIST-----IRDPEK-----DLKKL1-----KKKPREVDDFT-----SE 79	DE PF11_0392	Hypothetical protein.
Db	750 EK--DRNITNGNNNLYVEYNNSCIPPKMIPDGVRNSINKLNNTNTQRTSSVSY 806	GN OS OC NCBI_TaxID:36329; [1]	Plasmodium falciparum (isolate 3D7); Alveolata; Apicomplexa; Haemosporida; Plasmodium.
Qy	80 TGRMEEVDY-----KYYDKGNLIAYDDGTDIEYETKLDTEKSKTGYVULSPSK 128	RN RP SEQUENCE FROM N_A.	TRAIN=3D7;
Db	807 TNKNDIDENSFDMDPTINGRESKYISNNNNI--NGNSIGFNSSKLNDYTHONSMNV----- 858	RX MEDLINE=22555705; PubMed=12368864;	RX Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W.,
Qy	129 DGHEFILGKISNSVSKAAVYGNINYKSIEIKATKDYDFFSKTMF-----DIYANIN-- 179	RA	

Db	1092	HLITNMNKNNTLKEBLKNNNHYEVKK--NINE---DICKNDVANLGTDSTKQNFGVT	1145	Qy	8	SKFKNLQNGKEGSLLKKDTGVEHHHOENEESIKE--KSSFTIDRANISTIRDFFEN-KDLKK	64
Qy	254	YS-PSEKQOYLKDNNIIRKGY---ALKVTTYNPGKTDMEGN---GVSKEDIAKI	303	Db	728	SFDFN--DOEDKNCKD---KHSNDTTEKEHHRKNIDIVQKONSINSKEFGVKSFDFK	781
Db	1146	PSLDSTKQNFGVTPSLDSLKQNFGVTPSLDSLKQNFGVTHNLDSKRN	1205	Qy	65	L1-----KK-----	84
Qy	304	Q-KANPNLRLSETTIYASRNEYEDGRSTQSV--LMSALDGF----NIIRY---	347	Db	782	VYNNMKRYSINHKTKRNWLSEYCENDENMSKEIKMEEINTNMKNIKMNINNINYS	841
Db	1206	NFGVTPNLDSTKRN--FGVTHNLDSLKQNFGVTPSLDSLKQNFGVTPSLDSLKFDSEK	1263	Qy	85	-BEYDKYDDKGNI IAYD-----DGTLVETEKLDIBKSKYGVULPSKDGHF	132
Qy	348	-QVFTFKNDKGAEIKDQGLNLYTDSKVLVF---GKDKEYTEDKFVNVAEKEDGSM	402	Db	842	YDYDYDGEKSNLVSYKESAKSOKYASDLEKKHTKONIRGYLASKLKDHD-	900
Db	1264	EQYINKTLNNTFTNTKESENKPKTBEEKKVNIISLDPKNDKDNINNN NYERIKDGYIL	1322	Qy	133	EFLGKISNVSKNAKY-----YGNY-----	164
Qy	403	-----PDTKPVNLS-----MDKYNFNPSSKSNSKIIYVRNPEFYLRG	437	Db	901	BEKQKCKKIKKCKENLHNHYDKNKYDNYTNTDSNT1PFPYGIKSKNLISSYD	960
Db	1323	RNYRTNEEMEQKNNNSPLQNQININNSINNKHMDEGLIHNNNNNKYYNDESKYTFI	1382	Qy	165	FHSKTMFDLYANINDIVDGLAFAGDMRLFVKNDQKAEIKIRMPPEKIKETKSEPYVS	224
Qy	438	K1SD-KGGFNWELRVNNEVSDVNTLYIYGLDHIIDTRDFNIKLAVKGD1MDWGMKDYKANG	496	Db	961	-----DLVYNIN-----	975
Db	1383	NIEEGKQDNEKEHINENPVNKNPFI---HIEVK-----NKKCYTYTESPMKNNNNN	1432	Qy	225	SYGNVIBLGEGLSLSKRKPKPDNLTKMEEGKSYDSEKQOYLLKDNTI-----LRKGALKVT	279
Qy	497	FPPDKVTDMDGNNVYLQTGTSDF-LNAKAVGVHYQFLDNTKPEYNDPKGNTSTEYAD--G	552	Db	976	KNNNLILCDNKVYDHS DIMPSHKGKKCTDNNYVDFEFLNSGKGNLDDYLSNSGKGNY---	1032
Db	1433	IYD---NNNSNNYTYDNTTNVMGNKIKCCTKTYKNTLK--NRHKYNTNTDVNIRRG	1487	Qy	280	TYNPGKTDMLBENGIVSYKESDIAKIKQRANPNURALSETTIYASRNEYEDGRSTQSVLMMSA	339
Qy	553	KSVVP---NINDKRNNNGFGEIOPQH-----IYINGKEYTSFNDIKQ1DKTLNKI	601	Db	1033	NNNTYD-DDDDDSSTIHVDSDMDLKNYESISLUNNFSYTLD--NSKD-----I	1079
Db	1488	KN1KFVLNTNCYTN---VLKVTKHSMKKTKEVNLQKMY---DILKLUKKNKDINIKL	1539	Qy	340	DGFN1IIRYQVFETKMMNDKGAEIKDGNLYTDOSSKVLFLGKODKEYTEDKFVNVAEKEDG	399
Qy	602	WVDFARNTTVKREBFLNKTGEVSELSKHRVTVTQNG---KEMSTIVSBBDFFLPVV	657	Db	1080	DUNNNV--HNIVT-RINPKK---NDVNVNVEKGKINISLNSEEKEKRNEDQYKPLEEKKD	1131
Db	1540	LYNHL-----YLIKREEAHLLRVNTKVGVLINKGNHISRKQONNNIIKKRS1SSFS	1591	Qy	400	SMLFIDTKP---VNLSDMDKNTYFNPSSKSNSKIIYVRNPFYLQRGKISDKGGFTWLRVNESV	455
Qy	658	KGELEKGYQFDGWEISGP-----EGKKAQGYVINISKDTF1KPVFKK1EEKK	704	Db	1132	TIIFDENNDTNTVHNNDD--NNDTINSVHNNDD--NNNDNTIN-SVHVNND	1181
Db	1592	LNGQDNSN--DNSDVNSNNI1KNTIHYINDDIEKRNKSNYYQNVDDDDNSNMLKCEKK	1649	Qy	456	456 NYLITYGDLH--DNTRDFN1KLNVKGD1MDWGMKDYKANGFDPKVTDMDGNYLQ	512
Qy	705	EEENKPTDVSKKDDNPNQVNHQSOLNESRKEDLOREEHQSQSDSTRDVTATVLDKRN--	761	Db	1182	NNNDTNG-VHNNYDNNNDTINSVHVNND--HHKHTQSNNN--FSSSQENTYKNT	1233
Db	1650	NTKDKNTRKDNTKDNK-----TKDONTKDONTKDONTKDONTN1TKNNSYM	1701	Qy	513	GYSDLNAKAVGVHYOFYDNYKPEVNIDPKGKNT-----SIEYADGKSVVFN--IN	560
Qy	762	ISSKSTTNPNK 773		Db	1234	-----DNKTPCNYDKCSNEI1QDNDPNAKLYSSS1YSNTSBCN	1275
Db	1702	VSDNYILNDNK 1713		Qy	561	DKRNGFGEIOEQHYYINGKEYTSF--NDIKQ1DKTLNKIVVDFARN-----609	
				Db	1276	TKNYMGKKNKTKHLLNKHINDYYVDFSNELDKITKCKNS1KEKCDMKTNINNNNNNN	1335
				Qy	610	-----TVYKFLINKDGTGEVSELKPHRVVTIQLGKMSST--IVS8BDF-----652	
				Db	1336	EDKNRKRNKKNKQK1YINFSDGDIQ-----INNKCKNSNEKCKKINKENYICDET	1385
RESULT 12				Qy	653	-----ILPVYKGELEKGYQFDGWEISGFEGKDKADAGYVIN---LSKDTF1KPVPKIE	701
Q81E65		PRELIMINARY;	PRT; 3519 AA.	Db	1386	KNFTN1LIPKBYIMKHNYY-----SRKSFNNIKHQNGBDDEBTNT1KDEIKDKS	1434
AC				Qy	702	EKKEENKPTDPV-SKCKDNQVNHQSOLNESRKEDLOREEHQSQSDSTKOVATVLDK	760
DT	01-MAR-2003	(TREMBrel. 23, Created)		Db	1435	KHEEEVEYILLHKVDNKKNDQDDDHENDNNNNNNQ-----NDDH	1489
DT	01-MAR-2003	(TREMBrel. 23, Last sequence update)		Qy	761	NISSKSTTNPNK 773	
DT	01-MAR-2003	(TREMBrel. 23, Last annotation update)		Db	1490	PDNHNDNNNNK 1502	
DE		Hypothetical malaria antigen.					
GN	M26-32-10	OS Plasmodium falciparum (isolate 3D7).					
OC	Alveolata; Apicomplexa; Haemosporida; Plasmodium.						
RX	NCBI_TaxID=36329;						
RN	SEQUENCE FROM N.A.						
RA	Harris B., Lennard N., Clark L., Line A., Barron A., Corton C.,						
RA	Berriman M., Pain A., Hall N., Atkin R., Chillingworth C., Doggett J.,						
RA	Ormond D., Sanders M., Hayes R., Hall S., Quail M., Barrell B.,						
RA	Submitted (SEP-2002) to the EMBL/Genbank/DBJ databases.						
DR	EMBL; AL844509; CAD52401.1; -.						
KW	Hypothetical protein.						
SEQUENCE	3519 AA; 419077 MW; 15C47EFF2CO8CS393 CRC64;						
[1]	NCBI_TaxID=36329;						
RN	Query Match 6.2%; Score 251.5; DB 5; Length 3519;						
Best Local Similarity 18.9%; Pred. No. 0.058;							
Matches 173; Conservative 148; Mismatches 307; Indels 285; Gaps 40;							
RESULTS 13							
Q81I62							
ID	Q81I62						
AC	Q81I62;						
DT	01-MAR-2003 (TREMBrel. 23, Created)						
DT	01-MAR-2003 (TREMBrel. 23, Last sequence update)						
DT	01-MAR-2003 (TREMBrel. 23, Last annotation update)						
DE	Hypothetical protein.						

Search completed: January 28, 2004, 13:01:39
Job time : 48 secs